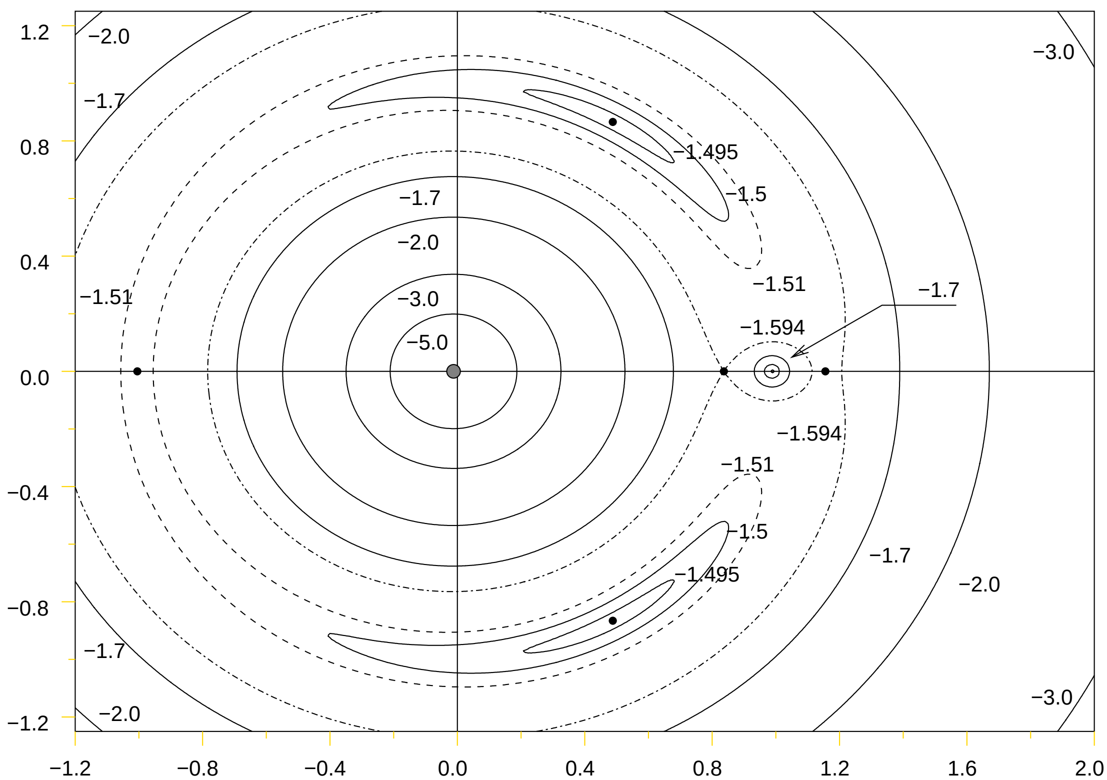
Introduction to Astronautics

Mechanics of Solar System Flight



by

David F. Chichka

June 23, 2010

The cover illustration is a sketch of the curves of zero velocity in the Earth-moon three-body problem.

Contents

List of Figures

Chapter 1

Introduction

Man has been trying to explain the motion of the planets (and the Sun and Moon, which were initially regarded in the same manner) for aeons. For most of this time, the attempt was to explain the apparent motion in terms of then-current theories. In Ptolemy's time, for example, the circle was considered a perfect shape, and it was assumed that the heavens were expressions of perfection. Therefore, it seemed clear that the motion of the heavenly bodies must be basically circular. When this failed to accord with observation, Ptolemy revised the simple scheme of nested spheres to include epicycles. The natural assumption that the Earth was the center of the motion was also slightly modified to allow for a center of rotation slightly removed from the center of the Earth.

In the sixteenth century (1543), Nicolaus Copernicus noticed that all the epicycles of the ptolemaic system had periods of exactly one year, and this led him to postulate that the Sun, rather than the Earth, was at the center. He still thought that orbits should be circular, and thus he still required small epicycles to explain retrograde motion. Still, he achieved a great reduction in the complexity of the description.

In the late sixteenth century, Johannes Kepler met Tycho Brahe. Brahe was a meticulous astronomer; Kepler a sickly mathematician. Kepler took observational data provided by Brahe and, for the first time, attempted to derive a theory to fit the data, rather than force the data to fit a philosophical assumption. Kepler published the first two of his three laws of planetary motion in 1609; the third followed nine years later.

Kepler's three laws will keep; we will not need to spend years fitting curves to data to derive them. Instead, they will become statements of the results of two statements made in the 1660's by Isaac Newton.

1.1 Newton

Just about the entirety of basic orbital mechanics can be derived from two statements made by Isaac Newton. The first of these is his second law of motion:

> The rate of change of momentum of a body (or system) is proportional to the impressed force, and is in the direction of the force.

The mathematical formulation of this law is

$$\mathbf{F} = \frac{d}{dt}(m\mathbf{v})$$

For our purposes, mass will almost always be fixed, so that we have instead the famous statement

$$\mathbf{F} = m\mathbf{a}$$

The second basic principle is Newton's law of universal gravitation. This might be expressed in words as

> Any two bodies experience an attraction towards each other. This attractive force is proportional to the product of the masses of the two bodies and inversely proportional to the square of the distance between them.

This is a vector equation, and can be written

$$\mathbf{F}_{1,2} = \frac{G m_1 m_2 (\mathbf{r}_2 - \mathbf{r}_1)}{\|\mathbf{r}_2 - \mathbf{r}_1\|^3}$$

where $\mathbf{F}_{1,2}$ is read, "the force on the first body (that is, the body labeled '1') due to the second body". The vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ are the positions of the two bodies, and m_1 and m_2 their masses. G is the *universal gravitational constant*, defined in an appropriate set of units.

This is all we need to know to derive Kepler's laws, *if* we use the knowledge correctly.

1.2 Reference Frames

Newton formulated his laws of motion with respect to an *inertial* frame. Such a frame must be non-accelerating, which requires both that the origin of the frame be moving at constant velocity and that the frame not be rotating. Such a frame is now known not to exist, and we settle for one that is *inertial enough*; it doesn't move enough, quickly enough, to measurably affect the outcome of our calculations. For throwing a baseball, a frame fixed in the Earth is inertial for all intents and purposes, and in fact the path of the ball covers so short a distance in comparison to the circumfrence of the Earth that we consider the Earth to be flat.

When firing long-range artillery, the assumption of a fixed Earth is no longer sufficient. In this case, the effect of the rotation of the Earth, and of the different radius from the axis of rotation to the surface at the firing point and at the target, is too large to ignore. For such problems, we include the rotation of the Earth, but ignore its motion through space.

What frame we choose for our problems will depend upon what we are studying. For orbital problems, the effect of a spinning planet will be negligible; for problems concerning rocket launch, it certainly will not.

1.3 A Few Notational Comments

Most of the notation of this book should be obvious. All variables will be described upon being introduced.

1.3.1 Vectors

In general, a vector quantity will be shown in the text and equations as a bold lowercase latin character, as was done in the last section. Matrices will be capitalized and not bold. Lowercase letters that are not bold will denote scalar quantities, as will Greek letters nearly always.

Throughout this book, vectors will be assumed to be in either two or three dimensions, depending on whether we are discussing planar motion or three-dimensional motion. In the few cases where we might need more general vectors, it will be made explicit. The magnitude of a vector is generally referred to by the lowercase non-bold letter, as

$$a = \|\mathbf{a}\|$$

The components of a vector in general will be denoted as a_1, a_2, a_3. When each component is associated with a specific unit vector, say $\hat{\imath}, \hat{\jmath}, \hat{k}$, they may be shown instead as a_i, a_j, a_k. A brief review of vectors in included in Appendix B.

Finally, the position vector $\mathbf{r}$ will generally be assumed to have the elements x, y, and z. This will not always be enforced, particularly when the use of several values would require awkward subscripting.

1.3.2 Astronomical symbols

This is a book about spacecraft and space travel, and will be concerned mostly with travel inside our solar system. When specific bodies within the solar system are used, subscripts denoting these bodies will be taken from traditional astrological usage. These symbols are

the Sun	☉		
the Moon	☾	Jupiter	♃
Mercury	☿	Saturn	♄
Venus	♀	Uranus	⛢
Earth	$\oplus$	Neptune	♆
Mars	♂	Pluto	♇

Examples in this book are worked out using the physical constants as listed in the appendices. Usually, the values used will be given in the examples themselves. While answers and preliminary results will only be listed to several significant digits, all calculations are carried out in extended precision, and intermediate results are used to the full precision to which they are calculated.

A Comment on Nomenclature: As celestial mechanics is a very old field of study, it might be expected that the terminology and symbols used are well-defined. That is not the case. Many different symbols are used, and many times different terms are used, for the various quantities involved in orbital mechanics. When there appears to be a standard, such as the use of $\mathbf{h}$ for the angular momentum vector, we have attempted to adhere to it. However, in the absence of such standards, we have not hesitated to assign notation of our own choosing.

Chapter 2

The Two-Body Problem

We will examine the simplest problem of orbital mechanics, that of two bodies in space, with no forces other than their mutual gravitational attraction. Due to the huge distances involved in orbital mechanics, this is in fact a good approximation of most planets in their motion about their stars, and of most moons about their planets. The effects of other planets and bodies in the same solar system are minor.

2.1 The General Equation of Motion

Consider the situation as shown in Figure 2.1. The two masses shown are considered to have negligible diameter when compared to the distance between them, and thus will be treated as points. The X, Y, Z frame is the inertial frame, in which Newton's equations are assumed to hold. The vectors $\mathbf{r}_1, \mathbf{r}_2$, and $\mathbf{r}_c$ are defined in this frame.

The vector $\mathbf{r}$, which gives the position of mass 2 with respect to mass 1, is defined in the local frame x, y, z. The x, y, z axes are parallel to their counterparts in the inertial frame, but move along with mass 1.

From basic mechanics, we find the point c, the center of mass of the system, to be

$$\mathbf{r}_c = \mathbf{r}_1 + \frac{m_2 \mathbf{r}}{m_1 + m_2} = \mathbf{r}_2 - \frac{m_1 \mathbf{r}}{m_1 + m_2} \tag{2.1}$$

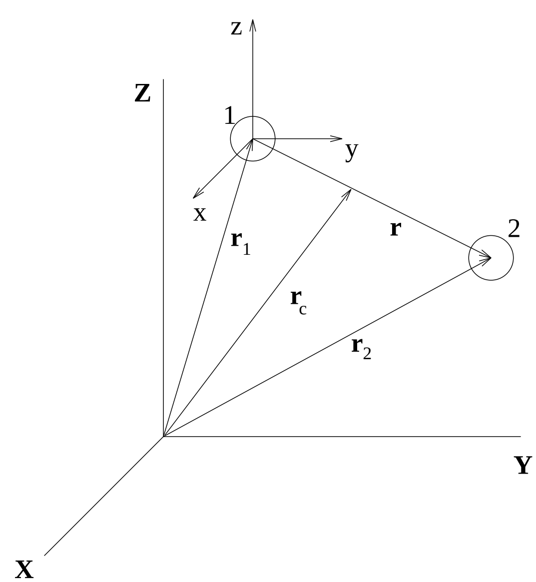


Figure 2.1: A two-body system.

From Newton's law of gravitation, we have the forces on the bodies as

$$\mathbf{F}_1 = \frac{Gm_1m_2\mathbf{r}}{r^3} \qquad \mathbf{F}_2 = -\frac{Gm_1m_2\mathbf{r}}{r^3} \tag{2.2}$$

Acceleration is the second derivative of the *inertial* position of a body. Since $\mathbf{r}_1$ and $\mathbf{r}_2$ are in the inertial frame, we can use Newton's second law and eqn. (2.2) to write

$$m_1\ddot{\mathbf{r}}_1 = \frac{Gm_1m_2\mathbf{r}}{r^3} \qquad m_2\ddot{\mathbf{r}}_2 = -\frac{Gm_1m_2\mathbf{r}}{r^3} \tag{2.3}$$

With this and eqn. (2.1), we can write

$$\ddot{\mathbf{r}}_1 = \ddot{\mathbf{r}}_c - \frac{m_2\ddot{\mathbf{r}}}{m_1+m_2}$$

with a similar result for $\mathbf{r}_2$. Using this, we substitute for $\ddot{\mathbf{r}}_1$ and $\ddot{\mathbf{r}}_2$ in (2.3) to get

$$m_1\left[\ddot{\mathbf{r}}_c - \frac{m_2\ddot{\mathbf{r}}}{m_1+m_2}\right] = \frac{Gm_1m_2\mathbf{r}}{r^3}$$

$$m_2\left[\ddot{\mathbf{r}}_c + \frac{m_1\ddot{\mathbf{r}}}{m_1+m_2}\right] = -\frac{Gm_1m_2\mathbf{r}}{r^3}$$

Adding these equations gives

$$(m_1+m_2)\ddot{\mathbf{r}}_c = 0$$
$$\implies \quad \ddot{\mathbf{r}}_c = 0$$

This means that the point c moves in a straight line, which depends only on the initial conditions. This is what we expect from applying Newton's first law to the overall system.

Since we assume we know the initial conditions, we will always have $\mathbf{r}_c$. We know $\mathbf{r}_1$ and $\mathbf{r}_2$ in terms of $\mathbf{r}_c, \mathbf{r}$, and the known values of m_1, m_2, and G. Thus, if we knew $\mathbf{r}$, we would have complete knowledge of the system.

We know that $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. From our two equations of motion (2.3), we can write

$$\begin{aligned} m_1m_2\left(\ddot{\mathbf{r}}_2 - \ddot{\mathbf{r}}_1\right) &= \frac{Gm_1^2m_2\mathbf{r}}{r^3} - \frac{Gm_1M_2^2\mathbf{r}}{r^3} = \frac{Gm_1m_2(m_1+m_2)\mathbf{r}}{r^3} \\ &\implies \quad \ddot{\mathbf{r}} = -\frac{G(m_1+m_2)\mathbf{r}}{r^3} \end{aligned}$$

It is convenient to use the notation $\mu = G(m_1+m_2)$. Thus the equation of motion becomes

$$\ddot{\mathbf{r}} = -\frac{\mu\mathbf{r}}{r^3} \tag{2.4}$$

Note: In practice, one mass is almost always vastly larger than the other, as is the case for a planet orbiting the Sun, or (even more so), a spacecraft orbiting a planet. Therefore in such cases it is general practice to ignore the smaller mass in the computation of μ. In the case of an orbit about the Earth, we would thus ignore the mass of the spacecraft, and write

$$\ddot{\mathbf{r}} = -\mu_{\oplus}\frac{\mathbf{r}}{r^3}$$

where $\mu_{\oplus} = Gm_{\oplus}$, that is, the universal gravitational constant times the mass of the Earth. The value of μ associated with a particular celestial body is called the *gravitational parameter* of the body. ◊

2.2 Solution of the Equation of Motion

Now that we have an equation for the motion of the second body about the first, we proceed to solve it.[1] This is a somewhat lengthy and intricate process, and we will proceed via three intermediate steps. Each of the steps is of great interest in its own right, as each involves a constant of the orbital motion. These constants will be used later to describe both the orbit and the motion of the body on the orbit.

In what follows, we will be using the velocity vector $\mathbf{v}$ interchangeably with $\dot{\mathbf{r}}$. The velocity is defined as the time rate of change of position, of course, so these quantities are equal by definition. The only reason to choose one over the other in a particular instance is to aid the clarity of the presentation.

2.2.1 Specific Energy

Because there are no dissipative terms in the equation of motion, we know that the total energy of the system is constant. We therefore look for some expression of the total energy. We recall that (see Appendix B)

$$\ddot{\mathbf{r}} \cdot \dot{\mathbf{r}} = \frac{1}{2}\frac{d}{dt}\left(\dot{\mathbf{r}} \cdot \dot{\mathbf{r}}\right) = \frac{1}{2}\frac{d}{dt}\left(v^2\right)$$

and note that this is the form obtained by taking the dot product of the left side of (2.4) with $\dot{\mathbf{r}}$. Now, $v^2/2$ is the per-unit-mass expression of the kinetic energy of a particle. Hence, we see that we can obtain an expression for the derivative of the kinetic energy in a straightforward manner.

[1]By "solve", we usually mean that we will attempt to find an explicit function $\mathbf{r}(t)$ that will, when differentiated, satisfy the equation. We will not find such a function, and o we are using the term rather loosely here. We will, however, be able to fully classify all types of motion that satisfy eqn. (2.4), so it isn't a complete misuse of the term.

Now, since $\mathbf{v}$ is expressed in an accelerating reference frame, it is not the inertial velocity of either particle. Still, this seems a promising avenue, so we take it, forming the dot product of equation (2.4) with $\dot{\mathbf{r}}$. We get

$$\ddot{\mathbf{r}} \cdot \dot{\mathbf{r}} = -\frac{\mu}{r^3}\mathbf{r} \cdot \dot{\mathbf{r}} \tag{2.5}$$

Recalling that

$$\dot{\mathbf{r}} \cdot \mathbf{r} = \dot{r}r$$

we can re-write (2.5) as

$$\frac{1}{2}\frac{d}{dt}\left(v^2\right) = -\frac{\mu\dot{r}}{r^2} \tag{2.6}$$

Now,

$$\frac{d}{dt}\left(-\frac{1}{r}\right) = \frac{\dot{r}}{r^2}$$

so (2.6) becomes

$$\frac{1}{2}\frac{d}{dt}\left(v^2\right) = \mu\frac{d}{dt}\left(\frac{1}{r}\right)$$

Finally, we can integrate this result to get

$$\frac{1}{2}v^2 = \frac{\mu}{r} + C \tag{2.7}$$

where C is a constant of integration. We refer to C as the *Specific Mechanical Energy* of the system, and re-write (2.7) to get

$$C = \frac{v^2}{2} - \frac{\mu}{r} \tag{2.8}$$

This quantity is often referred to simply as the energy of the orbit. Equation (2.8), in various forms, will be an important tool in describing the motion of bodies in orbit.

Recall that the kinetic energy of a body in motion is $mv^2/2$. The term "specific" is used here to denote that the quantity is expressed per unit of mass, so that the m drops out. The first term in the expression for C above is therefore clearly related to the kinetic energy of the body in orbit.

It is somewhat more difficult to see the relationship between the the second term and the potential energy of the body. We will return to this point later in the chapter.

2.2.2 Angular Momentum

The next major step is to cross the vector $\mathbf{r}$ into equation (2.4). We get

$$\mathbf{r} \times \ddot{\mathbf{r}} + \frac{\mu}{r^3}\mathbf{r} \times \mathbf{r} = 0$$

Since any vector crossed into itself produces zero, this reduces to

$$\mathbf{r} \times \ddot{\mathbf{r}} = 0 \tag{2.9}$$

Note that

$$\frac{d}{dt}(\mathbf{r} \times \dot{\mathbf{r}}) = \mathbf{r} \times \ddot{\mathbf{r}}$$

Using this, equation (2.9) becomes

$$\frac{d}{dt}(\mathbf{r} \times \mathbf{v}) = 0$$

Integrating this gives a vector constant of integration on the right hand side. Calling this constant vector $\mathbf{h}$, we have

$$\mathbf{r} \times \mathbf{v} = \mathbf{h} = \text{constant}. \tag{2.10}$$

This vector constant of integration is called the *specific angular momentum*. Note the similarity to the usual definition of angular momentum as $m\mathbf{r} \times \mathbf{v}$. Again, we are simply dropping the mass term.

Note that $\mathbf{h}$ being constant requires that the motion be restricted to a plane, because $\mathbf{r}$ is always normal to $\mathbf{h}$. Physically, we can see that the velocity is always in the plane to which $\mathbf{h}$ is normal; thus, the body can never move out of that plane. The plane of motion is referred to as the *orbital plane.*

2.2.3 The Laplace Vector

The third intermediate step in solving the equation of motion is to cross the original equation into $\mathbf{h}$. The result is

$$\ddot{\mathbf{r}} \times \mathbf{h} = \frac{\mu}{r^3}\mathbf{h} \times \mathbf{r}$$

Note that

$$\frac{d}{dt}(\dot{\mathbf{r}} \times \mathbf{h}) = \ddot{\mathbf{r}} \times \mathbf{h} + \mathbf{r} \times \dot{\mathbf{h}} = \ddot{\mathbf{r}} \times \mathbf{h}$$

(because $\mathbf{h}$ is a constant, so $\dot{\mathbf{h}}$ is of necessity zero). Using some results for vector triple products (see Appendix B) along with equation (2.10), we can also write

$$\mathbf{h} \times \mathbf{r} = (\mathbf{r} \times \mathbf{h}) \times \mathbf{r} = [\mathbf{v}(\mathbf{r} \cdot \mathbf{r}) - \mathbf{r}(\mathbf{r} \cdot \mathbf{v})] = r^2\mathbf{v} - (r\dot{r})\mathbf{r}$$

Now,

$$\frac{d}{dt}\left(\frac{\mathbf{r}}{r}\right) = \frac{1}{r}\mathbf{v} - \frac{\dot{r}}{r^2}\mathbf{r}$$

so our equation becomes

$$\frac{d}{dt}(\dot{\mathbf{r}} \times \mathbf{h}) = \mu\frac{d}{dt}\left(\frac{\mathbf{r}}{r}\right)$$

Since each side of the equation is a pure time derivative, the obvious thing to do is integrate, and so we do, getting

$$\dot{\mathbf{r}} \times \mathbf{h} = \mu \frac{\mathbf{r}}{r} + \mathbf{A} \tag{2.11}$$

$\mathbf{A}$ is another vector constant of integration, known usually as the *Laplace vector*.[2] We re-arrange (2.11) to define it as

$$\mathbf{A} = \mathbf{v} \times (\mathbf{r} \times \mathbf{v}) - \mu \mathbf{r}/r \tag{2.12}$$

The Laplace vector can be zero.

2.2.4 The Orbit Equation

So far, we have derived 7 constants of the motion for the body in motion (because $\mathbf{h}$ and $\mathbf{A}$ are three-dimensional vectors, they count for three scalar constants each). Due to some relationships between these constants, they actually only provide five pieces of information. This information is sufficient to define the size, shape, and orientation of the orbit, as will be shown in the next few chapters. However, they do not offer an intuitive feel for the orbit.

To get a better idea of the orbit, we turn equation (2.11) into a scalar equation by dotting it with $\mathbf{r}$ as follows:

$$\mathbf{r} \cdot \dot{\mathbf{r}} \times \mathbf{h} = \mu \frac{\mathbf{r} \cdot \mathbf{r}}{r} + \mathbf{A} \cdot \mathbf{r} \tag{2.13}$$

Using some triple-product identities from Appendix B, we have that

$$\mathbf{r} \cdot \dot{\mathbf{r}} \times \mathbf{h} = \mathbf{r} \times \dot{\mathbf{r}} \cdot \mathbf{h} = \mathbf{h} \cdot \mathbf{h} = h^2$$

and recalling the definition of the dot product, this gives us that

$$h^2 = \mu r + rA \cos \nu \tag{2.14}$$

where ν is the included angle between the Laplace vector and the radial vector $\mathbf{r}$. We re-write this equation as

$$r = \frac{h^2/\mu}{1 + (A/\mu) \cos \nu} \tag{2.15}$$

This equation is of the form

$$r = \frac{p}{1 + e \cos \nu} \tag{2.16}$$

[2] The Laplace vector is also associated with the name Runge-Lenz vector, and is referred to by Battin [2] as the eccentricity vector. Some authors go so far as to call it the Laplace-Runge-Lenz vector, but that's just inconvenient.

where $p = h^2/\mu$ and $e = (A/\mu)$. We will generally refer to this as the *orbit equation*.[3] This is the equation of the class of curves known as *conic sections* in mathematics, described in the polar coordinates r and ν.

The angle ν is known in orbital mechanics as the *true anomaly*. The term "anomaly" is often used in classical mathematics to describe an angle; the specification "true" serves to differentiate this angle from others that we will see in later chapters. The value of r is simply the distance between the two bodies.

2.3 Relative and Inertial Motion

Before delving into a discussion of the different kinds of orbits, it is convenient at this point to clear up one or two possible points of confusion.

2.3.1 Potential Energy

We said that eqn. (2.8) is the equation for the specific mechanical energy. In other words

$$C = T + U = v^2/2 - \mu/r$$

where T stands for kinetic energy, and U for the potential energy.[4] Leaving aside the kinetic energy for now, consider the potential energy term.

We define the potential energy in a conservative force field as the work done by the conservative force as a body moves from some arbitrary position to the reference position (we state without further explanation that gravitation is a conservative field). At some point $\mathbf{r}$, the work done in moving a unit mass to $\mathbf{r} + d\mathbf{r}$ is

$$dW = \mathbf{F} \cdot d\mathbf{r} = -\frac{\mu \mathbf{r} \cdot d\mathbf{r}}{r^3} = -\frac{\mu}{r^2} dr$$

Integrating from r to the reference point R, we get

$$W_{r \to R} = -\mu \int_r^R \frac{dr}{r^2} = \mu \left(\frac{1}{R} - \frac{1}{r} \right)$$

Now, the reference point is arbitrary. We choose it to be at infinity, so

$$U(r) = \lim_{R \to \infty} W_{r \to R} = -\mu/r$$

[3] It is also known in some texts as the trajectory equation, and many give it no name at all. We use the term here for convenience of reference only.

[4] Many texts use a negative sign here. The classic Brouwer and Clemence [4], for example, state the energy equation as $C = T - U$ (p. 27) (their symbols are actually different; we have converted to our notation for clarity). The difference comes in the definition of potential energy as the negative of a function known as the potential function; U is then the potential function. See Thomson [15] for more details of the argument presented here; see [4] for the opposite approach.

and we have the potential as in eqn. (2.8).

Note that while the we may choose R freely for the purpose of this exercise, it is not truly arbitrary in the derivation of eqn. (2.8). This is because the value of a constant of integration in general is chosen to meet the boundary conditions of the particular problem being evaluated. Thus the way eqn. (2.8) is derived tells us that C is constant; eqn. (2.8) itself tells us how to evaluate C to meet the conditions of a particular problem. The exercise just completed demonstrates only that $-\mu/r$ is indeed the potential energy, if $R \to \infty$ is taken as the reference.

2.3.2 Kinetic Energy

In expressing the energy of the system, we used the relative velocity $\mathbf{v}$, which we noted is expressed in an accelerating frame. By this we mean that it is the velocity of the second body relative to the first, and the first body is itself accelerating. In contrast, the conservation of energy is generally expressed relative to an inertial frame.

Consider the total kinetic energy T_I of the system, expressed in terms of the inertial velocities $\mathbf{v}_1$ and $\mathbf{v}_2$. We have

$$T_I = \frac{m_1}{2}\mathbf{v}_1 \cdot \mathbf{v}_1 + \frac{m_2}{2}\mathbf{v}_2 \cdot \mathbf{v}_2. \tag{2.17}$$

Recalling the definition of the center of mass (2.1), we have that

$$\mathbf{v}_1 = \mathbf{v}_c - \frac{m_2\mathbf{v}}{m_1 + m_2} \qquad \text{and} \qquad \mathbf{v}_2 = \mathbf{v}_c + \frac{m_1\mathbf{v}}{m_1 + m_2}.$$

Substitution of these expressions into (2.17) and multiplying out gives the kinetic energy to be

$$T_I = \frac{m_1 + m_2}{2}\mathbf{v}_c \cdot \mathbf{v}_c + \frac{1}{2}\left(\frac{m_1 m_2}{m_1 + m_2}\right)\mathbf{v} \cdot \mathbf{v}.$$

We have already seen that the velocity of the enter of mass is constant. Thus the inertial kinetic energy is a constant added to the relative kinetic energy.

The energy C is therefore not quite the total mechanical energy as seen from the inertial point of view. It is clearly intimately related, however, and is clearly a conserved quantity. The second term, $-\mu/r$, is genuinely the potential energy, and the sum of the two terms is a conserved quantity. It will do no harm to consider C to be the total mechanical energy of the system.

2.4 Properties of Conic Sections

We have said that equation (2.16) is the equation for a conic section. Conic sections are the curves that result when a plane is passed through a pair of infinite circular cones sharing a vertex and having the same axis of symmetry. The lines of intersection draw the conic section in space.

Another common way that conic sections are defined, especially in geometry texts, is

> **Definition:** A conic section is either a circle or a locus of points such that for each point, the ratio of its distance from a given point (the *focus*) to its distance from a given line (the *directrix*) is constant. This constant is the *eccentricity* e.

This definition is true, but for us is not terribly useful.

Recall equation (2.16).The conic sections are usually thought of as falling into four basic categories, depending upon the value of the eccentricity e. These are:

$$\begin{array}{rl} e = 0 & \text{circle} \\ 0 < e < 1 & \text{ellipse} \\ e = 1 & \text{parabola} \\ e > 1 & \text{hyperbola} \end{array}$$

We will mostly be concerned with the ellipse and the hyperbola. Circles are special cases of the ellipse, where many computations simplify because $e = 0$. The parabola is another special case, and can be thought of as a limiting case of both the ellipse (as the eccentricity rises to 1), and the hyperbola (as e again approaches 1, this time from above).

Clearly, a plane passing only through the vertex would create a point, and a plane tangent to the side of the cones would create a line. These are known as *degenerate* conic sections, and except for a brief discussion in a later section, we will for the most part ignore them. There is also the possibility of the eccentricity being negative; some authors describe an ellipse for which $e < 0$ as a *subcircular* ellipse. Such a condition does not arise in orbital mechanics.

2.4.1 Terminology

The following list includes the common terms describing the geometry of conic sections. Not all of them are meaningful for all types of conics.

Focus: All conic sections have at least one *focus*; the ellipse and the hyperbola have two. In these cases, one is known as the *empty focus*. The circle is a special case of the ellipse, and the focus and empty focus are both at the center.

Physically, the focus is the location of the first body, around which the second body orbits. In the mathematical derivation of this chapter, the focus is the point from which the vector $\mathbf{r}$ is drawn.

Axes: Conic sections have a major and a minor axis. The major axis is the line that runs through both foci; the minor axis is perpendicular to it and crosses it halfway between the two foci. The axes are not uniquely defined for the circle, and for any non-elliptic section the minor axis has little meaning.

Latus Rectum: The width of each curve *at the focus* is the latus rectum; half of this distance is the semi-latus rectum, usually denoted p.

Apses: The points on the curve at which it crosses the major axis are called *apses*. These points occur by definition when the value of the true anomaly ν equals zero and π. The first of these is called the *periapsis*, and the second the *apoapsis*.[5]

semimajor axis: The *semimajor axis* a is half of the distance between the apses of a conic section. For a parabola, for which the empty focus is taken to be at infinity and which thus has only one apse, $a \to \infty$. For a hyperbola, a is taken to be negative.

semiminor axis: The *semiminor axis* b is well-defined only for the ellipse; it is the distance from the point at which the axes cross to the point at which the minor axis crosses the ellipse.

c: The distance between the foci is $2c$, but there is no name given to this distance in most orbital texts.

In the next few sections, we will take a look at the different conic sections, beginning with the ellipse. For the most part, relations defined for the ellipse will hold true for other conic sections, when they make sense. Exceptions will be noted.

2.5 Elliptic Orbits

The ellipse is the only closed conic section (we will consider the circle as a special case of the ellipse), and so all closed orbits are ellipses. In this section, we will assume that $e > 0$, so that the ellipse is not a circle. The special qualities of the circle will be discussed later.

Recall the manner in which the trajectory equation was derived. We generally think of a small body orbiting a much larger one, for example the Earth orbiting the Sun. The trajectory equation describes the path of the bodies in relationship to the other, and thus the focus of the ellipse is at the center of mass of the body being considered the “stationary mass”; the one about which the other orbits. We made no actual assumptions about the

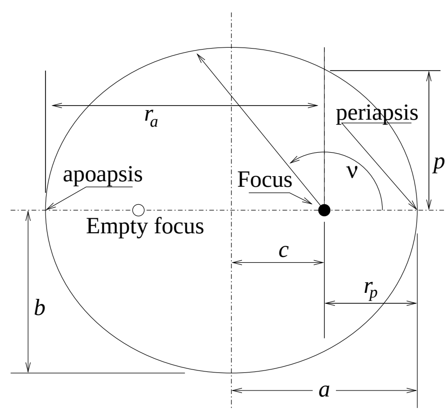


Figure 2.2: Variable definitions for the ellipse.

[5]The terms ’periapsis’ and ’apoapsis’ are often ignored in favor of *pericenter* and *apocenter*. This is a matter of taste.

relative masses, however, and mathematically it would be as valid to say that the Sun describes an elliptic orbit about the Earth. It will not be troublesome, however, to continue to think of the larger mass as stationary, and the smaller the orbiting mass.

Let r_p and r_a be the distances from the focus[6] to the periapsis and the apoapsis, respectively. It is easily enough seen that

$$r_p = \frac{p}{1+e} \qquad\qquad r_a = \frac{p}{1-e}$$

Since $r_p + r_a = 2a$, we have

$$\begin{aligned} 2a &= \frac{p}{1+e} + \frac{p}{1-e} \\ \implies &\quad 2a(1+e)(1-e) = (1-e)p + (1+e)p \\ \implies &\quad p = a(1-e^2) \end{aligned}$$

This result can be used to obtain simple expressions for the distance to the apses in terms of the semimajor axis as well:

$$r_p = \frac{p}{1+e} = \frac{a(1-e^2)}{1+e} = a(1-e) \tag{2.18}$$

This is true for all conic sections other than the parabola. The complementary statement that $r_a = a(1+e)$ is clearly true for the ellipse. r_a is the largest value r can achieve for an elliptic orbit.

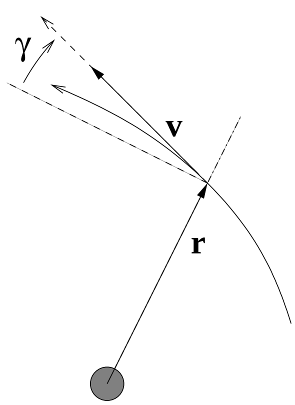


Figure 2.3: Flight path angle.

Recall that for us, $p = h^2/\mu$. The gravitational parameter $\mu = G(m_1+m_2)$ is determined by the physical properties of the system, so that for a specified pair of masses, the semi-latus rectum of a particular orbit is purely dependent upon the magnitude of the angular momentum of the orbit.

Recall the vector definition of $\mathbf{h}$:

$$\mathbf{h} = \mathbf{r} \times \mathbf{v}$$

This gives that $h = rv\sin\phi$, where ϕ is the included angle. Since $\mathbf{v}$ is normal to $\mathbf{r}$ at the apses, we have that

$$h = v_a r_a = v_p r_p \tag{2.19}$$

Consider Figure 2.3. By analogy with aircraft flight, we may call the line normal to the radial vector the *local horizontal*, and the angle γ between the local horizontal and the velocity vector the *flight path angle*. Again using the equation for the

[6]From here on, "focus" will mean "primary focus"; there is little use for the empty focus, and if it referred to at all, it will be explicitly denoted as empty.

magnitude of the cross product, we have at any point on the orbit

$$\begin{aligned} h &= rv\sin\phi = rv\sin\left(\frac{\pi}{2}-\gamma\right) \\ \Longrightarrow &\quad \cos\gamma = \frac{h}{rv} \end{aligned} \tag{2.20}$$

Because this is derived from the definition of **h**, it is true for all conic sections.

Now consider the value of the energy C for an elliptic orbit. At periapsis, we can write

$$C = \frac{v_p^2}{2} - \frac{\mu}{r_p}$$

Using the results of this section, we can re-write this as

$$C = \frac{h^2}{2r_p^2} - \frac{\mu}{r_p} = \frac{\mu p}{2r_p^2} - \frac{\mu}{r_p}$$

Writing p and r_p in terms of a and e produces

$$C = \frac{\mu a(1-e^2)}{2a^2(1-e)^2} - \frac{\mu}{a(1-e)}$$

and this expression simplifies to

$$C = -\frac{\mu}{2a} \tag{2.21}$$

An obvious result of this is that, since a is clearly positive for any ellipse the energy C of any elliptic orbit is negative. While negative energy may not seem to make much sense, it is due to the reference point for zero potential energy.

Example 2.1. A vehicle in orbit about the Earth has periapsis radius $r_p = 7000\,\text{km}$ and velocity $v_p = 9\frac{\text{km}}{\text{s}}$. Find the apoapsis radius and velocity.

Because the orbit is about the Earth, we use the gravitational parameter for the Earth. The value used in this text is $\mu_\oplus = 3.98601 \times 10^5\,\text{km}^3/\text{s}^2$. Note the need to use consistent units.

Begin by computing the orbital energy:

$$C = \frac{v^2}{2} - \frac{\mu}{r} = -16.443\,\text{km}^2/\text{s}^2$$

From eqn. (2.21) we then have

$$a = -\frac{\mu}{2C} = 12120.7\,\text{km}$$

We can now compute the eccentricity from the relation $r_p = a(1-e)$ to be

$$e = 1 - \frac{r_p}{a} = 0.42248$$

and from this the periapsis radius $r_a = a(1 + e) = 17241.4\,\text{km}$. Then because the angular momentum is constant, we have

$$r_p v_p = r_a v_a \quad \Longrightarrow \quad v_a = \frac{r_p v_p}{r_a} = 3.65540\,\frac{\text{km}}{\text{s}} \qquad \spadesuit$$

Note that there is almost never only one way to solve a problem in orbital mechanics. The method above is one approach to the problem posed. However, the problem could also have been solved by finding the angular momentum h, from that the semi-latus rectum, and then the eccentricity.

2.6 Hyperbolic Orbits

The other main type of conic section is the hyperbola. Like the ellipse, it is a symmetric orbit, but in this case, only mathematically. The two parts never meet.

2.6.1 Geometry of the Hyperbola

A hyperbola is shown in Figure 2.4. The solid line represents the actual orbit, and is curved about the primary focus. The broken line is the imaginary part of the orbit, and curves about the empty focus. The two parts of the orbit are mirror images, and the major axis is the line running through the two foci.

Because the eccentricity of a hyperbola is greater than one, there is some angle for which $1 - e\cos\nu = 0$. As that value is approached, the distance r goes to infinity. For physical orbits, this is all that is possible. Mathematically, however, the other part of the hyperbola does exist. One way to think of this is to imagine that the hyperbola extends to infinity and "wraps around", to come back in from negative infinity. It is not, however, necessary to have any picture in mind to make the equations work.

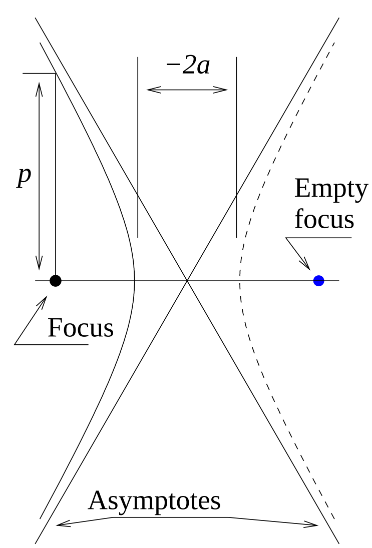


Figure 2.4: Terminology of the hyperbola.

From the orbit equation, it is clear that $r_p = p/(1 + e)$, for any conic section. Defining the apoapsis distance as the value of r when the true anomaly equals π, we again have $r_a = p/(1 - e)$, and in this case the value is negative. Looking at Figure 2.4, this is perfectly reasonable. Note that at $\nu = \pi$ the vector $\mathbf{r}$ would be running to the left in the sketch, while the point at which the imaginary arc of the hyperbola crosses the major axis is to the right of the primary focus, indicating that its position is negative in the **r** direction.

Because the equations are mathematically valid, we suppose that it is also still valid to write

$$2a = r_p + r_a$$

Then, following what we did before, once again $p = a(1 - e^2)$. Note that the semi-latus rectum p still makes perfect sense; the primary branch of the hyperbola will still pass above the focus and the distance will still be positive. We then have

$$a = \frac{p}{1 - e^2} < 0$$

It should be noted that this does not say that the absolute distance between the periapsis and apoapsis is negative. Recall that the logic for letting r_a be negative was that it was a location measured from the primary focus in a specified direction. This also applies to the semimajor axis. Think of it as the distance from the periapsis to the apoapsis, positive in the direction from the periapsis towards the primary focus.

There is no semiminor axis for the hyperbola. Of course, the minor axis is still defined, when thought of as the axis of symmetry between the two curves of the hyperbola. However, neither branch of the hyperbola ever intersects that axis, so b cannot be defined.

2.6.2 Hyperbolic Orbits

Now we apply the geometry of the hyperbola to the orbit equation, and vice-versa.

Once we accept that the semimajor axis a exists and is meaningful despite being negative, we can repeat the manipulations we did for the elliptic orbit and show that

$$C = -\mu/2a$$

Because a is negative, this means that the energy is positive on a hyperbolic orbit.

We have said that our reference point for potential energy is $R \to \infty$. This means that the potential energy as a function of the distance of the body in orbit from the central body is defined to be zero when the body is at an infinite distance. As a result, any body in orbit with a negative specific energy can never reach the reference point. That is, the value of r remains finite.

On the other hand, $C > 0$ implies that not only can the body reach infinity, but it will still have velocity when it gets there (because the potential energy is zero, the remainder is kinetic energy). Since the energy of the orbit is constant, we have

$$\frac{v^2}{2} = C + \frac{\mu}{r}$$

As $r \to \infty$, the last term goes to zero and

$$v_\infty^2 = 2C$$

which clearly makes no sense at all for $C < 0$, but is perfectly acceptable for hyperbolic orbits. The value v_∞ is known as the *hyperbolic excess speed*. It is the speed that a space probe has as it gets far from the solar system, for example.

The angle of the asymptotes is also easy to find. Since

$$r = \frac{p}{1 + e\cos\nu},$$

we have

$$r \to \infty \Longrightarrow 1 + e\cos\nu_\infty \to 0.$$

This gives

$$\cos\nu_\infty = -1/e. \tag{2.22}$$

Because $\cos(-\nu) = \cos(\nu)$, this tells us that value of the true anomaly on the primary branch of the hyperbola (which is also the physically realizable trajectory) is limited to $\pm\nu_\infty$. The values of ν such that $\cos\nu < -1/e$ correspond to the imaginary branch of the hyperbola.

Note also that because $\cos\nu_\infty < 0$ for all finite values of e, the true anomaly always sweeps at least through $-90°$ to $+90°$.

Example 2.2. What are the conditions at periapsis for a spacecraft on a trajectory that will approach infinity with $v_\infty = 2\,\mathrm{km/s}$ with $\nu_\infty = 120°$?

From the information available, we can easily compute the energy:

$$C = \frac{v_\infty^2}{2} = 2\,\mathrm{km}^2/\mathrm{s}^2$$

Compute the eccentricity from knowledge of ν_∞:

$$\cos\nu_\infty = -\frac{1}{e} \quad \Longrightarrow \quad e = \frac{-1}{\cos\nu_\infty} \quad \Longrightarrow \quad e = 2$$

Now, $a = -\mu_\oplus/2C = -99650\,\mathrm{km}$, so

$$r_p = a(1-e) = -a = 99650\,\mathrm{km}$$

The corresponding value of velocity can be computed from

$$\frac{v_p^2}{2} = C + \frac{\mu_\oplus}{r_p} = 2 + \frac{\mu_\oplus}{99650} = 6\frac{\mathrm{km}^2}{\mathrm{s}^2}$$

from which $v_p = \sqrt{12} \approx 3.4641\,\mathrm{km/s}$. ♠

2.7 Special Cases and Degenerate Conics

We will briefly discuss the special cases of circular and parabolic orbits. Circular orbits in particular are very important because many Earth satellite missions require them, and because we often approximate planetary orbits as circular. We will also mention the degenerate conics, though they almost never occur in astrodynamics work.

2.7.1 Circular Orbits

As mentioned, the circle is the special case of the ellipse for which the eccentricity is zero. This means that the orbit equation reduces to

$$r = p = a$$

The Laplace vector for a circular orbit has zero magnitude, and therefore there is nothing from which to measure the true anomaly (in practice, a reference is simply specified).

Because r is constant on a circle, the velocity is also constant on a circular orbit. The circular orbit velocity can be found from

$$C = \frac{v_c^2}{2} - \frac{\mu}{r}$$

From above, $r = a = -\mu/2C$, so this becomes

$$\frac{-\mu}{2r} = \frac{v_c^2}{2} - \frac{\mu}{r} \quad \Longrightarrow \quad v_c = \sqrt{\frac{\mu}{r}}$$

Because the velocity is constant, it is easy to compute the period of a circular orbit. The circumference is $2\pi r$, so the period P can be found from

$$P v_c = 2\pi r \quad \Longrightarrow \quad P = \frac{2\pi a^{3/2}}{\sqrt{\mu}}$$

We will see later that this expression is true for all elliptic orbits.

2.7.2 The Parabolic Orbit and Escape Speed

The parabola can be thought of as a transitional case between the elliptical and hyperbolic orbits. On the one hand, it is an ellipse for which $e \uparrow 1$. On the other, it is a hyperbola for which $e \downarrow 1$.[7]

[7] The symbols $\uparrow$ and $\downarrow$ here mean “approaches from below” and “approaches from above”, respectively.

In thinking of the parabola as the limiting case of an ellipse, we have the length of the ellipse becoming infinite. This is easily seen from

$$a = \frac{r_p}{1-e} \tag{2.23}$$

which is true for any conic section for which r, a, and e are defined. As $e \to 1$, the only way for this relation to be valid is for either r_p to become vanishingly small, or $a \to \infty$. The first possibility is nonsensical.

Since eqn. (2.23) holds for hyperbolas as well, and r_p is positive for hyperbolas, the same argument has $a \to -\infty$ as $e \downarrow 1$ for the hyperbola. Thus in the one case, the semimajor axis is going to positive infinity, and the apoapsis is then moving towards infinity. Then, as e crosses unity, the semimajor axis suddenly goes to negative infinity, and the apoapsis, now on the imaginary branch of the hyperbola, is moving back towards the periapsis, but from the negative direction.

In both cases, however, the energy is going to zero. Note that

$$\lim_{e \uparrow 1} C = \lim_{a \to \infty} \frac{-\mu}{2a} = 0$$

and

$$\lim_{e \downarrow 1} C = \lim_{a \to -\infty} \frac{-\mu}{2a} = 0$$

We therefore see that the specific energy for a parabolic orbit is zero.

Consider this in light of the hyperbolic excess speed. Since $v_\infty^2 = 2C$, this clearly indicates that the excess speed is zero. Thus a craft on a parabolic orbit will reach an infinite distance from the planet, but will have no energy or velocity left.

We might also consider this in light of the velocity at some specified orbital radius. Suppose that the spacecraft is at $r = r_0$, some specified value. From our previous discussion, it is clear that the parabola is the minimum energy orbit for which the spacecraft can attain an infinite distance. Since both r_0 and μ are constants, the energy equation $C = v^2/2 - \mu/r_0$ implies that C increase as v increases.

Let

$$v_e = \sqrt{\frac{2\mu}{r_0}}. \tag{2.24}$$

This will result in $C = 0$, so that $r \to \infty$ as the time of flight gets large. Any lower velocity at r_0 will result in a negative value for the specific energy, so that the distance can never become infinite, and the orbit is an ellipse. For this reason, the parabolic orbit is also known as an *escape orbit*, and v_e as defined in eqn. (2.24) is known as the *escape speed*.

Note that in practice, "infinite" means "very large". Since under any circumstances, infinite time would be required to attain infinite distance, we are never truly interested in doing so. However, the escape speed is a close approximation to the speed necessary to "escape" from the gravitational pull of a planet or star, and is used for initial approximations in interplanetary and interstellar mission analyses. These topics will be dealt with in detail in later chapters.

2.7.3 Degenerate Conics

The degenerate conics are the point and the line. The first of these is simply disposed of, while the second requires some discussion.

Point

The point is not a physically realizable orbit. It corresponds to the case in which the body in orbit is at the center of the central body, and has no velocity relative to it. More precisely, it means that the two masses m_1 and m_2 used in deriving the equation of motion are centered at the same point. Newton's law of gravitation is not defined for this case; it would consider the two bodies to be a single body. We will waste no further space on the point as a possible orbit.

Line orbits.

Recall the equation of motion

$$\ddot{\mathbf{r}} = -\frac{\mu \mathbf{r}}{r^3}.$$

This is defined (and correct) for all $r \neq 0$. However, in deriving the Laplace vector and then the orbit equation, we made the tacit assumption that the angular momentum was nonzero. All of the analysis in section 2.2.3 is still correct, but the Laplace vector comes out to be zero.

The line is the orbit resulting from $r > 0$ and the angular momentum $\mathbf{h} = 0$. In this case, we have that either $\mathbf{v} = 0$ or that $\mathbf{v}$ is aligned with $\mathbf{r}$ (so that the cross product is zero). In either case, we see that since $\ddot{\mathbf{r}}$ is always aligned with $\mathbf{r}$, there will never be a component of velocity that is not aligned with $\mathbf{r}$. Consequently, the velocity will always be either directly towards or away from the central body. There is no version of the orbit equation corresponding to this case.

Note that for this orbit, the energy and angular momentum relations still hold. The angular momentum is constant at zero, and the specific energy can still be computed as $C = v^2/2 - \mu/r$. The only point at which the energy equation breaks down is at $r = 0$, which can be treated mathematically as a singularity.[8] Physically, of course, there is no need to consider what happens as $r \to 0$, as the spacecraft (or comet, or asteroid, or planet, or...) will have crashed into the Earth (planet, Sun, ...) before that happens.

2.8 Kepler's Laws

In the late 16th century, the Danish astronomer and aristocrat Tycho Brahe made meticulous observations of the planets. When Brahe died in 1601, his notebooks

[8] We will not get into this here.

were inherited by Johannes Kepler (1571 – 1630), a sickly mathematician that he had chanced to meet about 18 months previously. Brahe appears to have been lacking the gifts of mathematical insight possessed by Kepler, and the latter had already begun working on the observational data. Finally, in 1609, Kepler set forth two laws of planetary motion, and ten years later, a third.

Kepler's laws of planetary motion are:

- The orbit of each planet is an ellipse, with the Sun at one focus.
- The line joining the planet to the Sun sweeps equal areas in equal time.
- The square of the period of the planet's orbit is proportional to the cube of its mean distance from the Sun.

These laws are purely descriptive. They make no attempt to explain *why* the planets behave as they do. They serve, however, as a remarkable example of rigorous reduction and analysis.

Since Kepler's laws incorporate several years of observations, it is necessary that any theory of planetary motion agree with them. To this end, we can somewhat verify the results of the two-body analysis if we can use these results to derive Kepler's laws. The first of them has already explicitly been derived, when the orbit equation turned out to be the equation of conic sections. In the following sections, we will derive the second and third of Kepler's laws.

2.8.1 Kepler's Second Law

The second law concerns the area swept by the line joining the Sun to the planet. The derivation requires first an expression for the rate at which this area is swept.

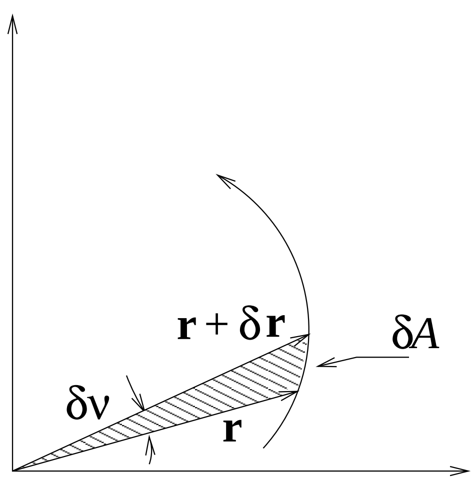


Figure 2.5: Area swept in time dt.

Consider Figure 2.5. In this figure, the shaded area is the area swept by the radius vector in an incremental time dt. Since dt is very small, the corresponding change in the anomaly ν is also small, and the curve from $\mathbf{r}$ to $\mathbf{r} + \delta\mathbf{r}$ is essentially a straight line segment. Using this approximation, the shaded area δA is given as

$$\delta A = \frac{1}{2}\left[r^2 \sin\delta\nu + r dr \sin\delta\nu\right]$$

where dr is the change in the magnitude of r (note that this is *not* the same as the magnitude of $\delta\mathbf{r}$). Since $\delta\nu$ is small,

$\sin\delta\nu \approx \delta\nu$. Also, we know that $dr \approx \dot{r}dt$ and $\delta\nu \approx \dot{\nu}dt$ Therefore, we have

$$\delta A = \frac{1}{2}\left[r^2\dot{\nu}dt + r\dot{r}\dot{\nu}dt^2\right]$$

The derivative of A with respect to time is by definition the limit

$$\frac{dA}{dt} = \lim_{dt\to 0}\frac{\delta A}{dt} = \lim_{dt\to 0}\left[\frac{1}{2}r^2\dot{\nu} + \frac{1}{2}r\dot{r}\dot{\nu}dt\right] = \frac{1}{2}r^2\dot{\nu} \tag{2.25}$$

We can now relate this result to the specific angular momentum. We have that

$$\mathbf{h} = \mathbf{r}\times\mathbf{v}$$

and therefore that

$$h = rv\sin\phi$$

where ϕ is the included angle.

Figure 2.6 shows this schematically. We can see from this figure that

$$v\sin\phi = r\dot{\nu}$$

and thus that

$$\frac{dA}{dt} = \frac{1}{2}r^2\dot{\nu} = \frac{r^2 v\sin\phi}{2r} = \frac{1}{2}h \tag{2.26}$$

Since we already know that h is constant, we have shown that dA/dt is constant throughout the orbit. Therefore, over any period, we have that

$$\Delta A = \frac{h}{2}\Delta t$$

which is a mathematical expression of Kepler's second law.

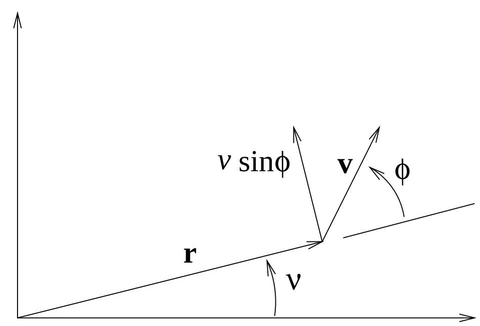


Figure 2.6: Relating $\mathbf{r}$, $\mathbf{v}$, and $\dot{\nu}$.

2.8.2 Kepler's Third Law

The third law concerns the period of the orbit and its relationship to the semimajor axis. In deriving this law, it helps to know in advance that the area of an ellipse is given by $A = \pi ab$.

If P is the period of an elliptic orbit, we have from (2.26) that

$$A\int_0^P \dot{A}dt = \int_0^P (h/2)dt = \frac{hP}{2} = \pi ab$$

The semiminor axis b is related to the semimajor axis by the relation

$$b = a\sqrt{1-e^2}$$

so that

$$P = \frac{2\pi ab}{h} = 2\pi\frac{a^2\sqrt{1-e^2}}{h}$$

Recall that $h = \sqrt{p\mu} = \sqrt{a\mu(1-\varepsilon^2)}$, which we plug in to give

$$P = \frac{2\pi a^2\sqrt{1-e^2}}{\sqrt{a\mu(1-e^2)}} = \frac{2\pi}{\sqrt{\mu}}a^{3/2}$$

This in turn implies

$$P^2 = \frac{4\pi^2}{\mu}a^3 \tag{2.27}$$

and (2.27) is a mathematical expression of the third law.

2.8.3 Measurement of Planetary Mass

As an implementation of the third law, consider the problem of measuring the mass of a planet that has a satellite. In this situation, there are two orbits to be considered. One is the orbit of the satellite about the planet, and the other the orbit of the planet about the Sun.

Let the subscript s denote variables relating to the satellite, and p those relating to the planet. Then by Kepler's third law, we can write

$$G(m_p + m_s) = 4\pi^2 a_s^3/P_s^2 \tag{2.28}$$

and

$$G(m_\odot + m_p) = 4\pi^2 a_p^3/P_p^2 \tag{2.29}$$

The values of the semimajor axis and the period of the orbits can be obtained to very high accuracy by from observations.

Dividing (2.28) by (2.29) gives

$$\frac{m_p + m_s}{m_\odot + m_p} = \left(\frac{a_s}{a_p}\right)^3 \left(\frac{P_p}{P_s}\right)^2$$

On the assumption that $m_\odot \gg m_p$ and $m_p \gg m_s$, we can then write

$$\frac{m_p}{m_\odot} = \left(\frac{a_s}{a_p}\right)^3 \left(\frac{P_p}{P_s}\right)^2 \tag{2.30}$$

which gives us the mass of the planet in terms of the mass of the Sun.

Example 2.3. Phobos travels about Mars in only 0.31891 sidereal days and the mean radius of its orbit is 9378 km.[9] The orbit of Mars about the Sun requires 688.9 sidereal days and has a mean radius of 227.8×10^6 km

Solution: From (2.30), we have

$$\frac{m_{♂}}{m_{\odot}} = \left(\frac{a_s}{a_{♂}}\right)^3 \left(\frac{P_{♂}}{P_s}\right) = \left(\frac{9378}{227.8 \times 10^6}\right)^3 \left(\frac{688.9}{0.31891}\right)$$

where the subscript s refers to Phobos (the satellite, in this case). Working out the fractions gives

$$\frac{m_{♂}}{m_{\odot}} = 3.256 \times 10^{-7} \Longrightarrow m_{\odot} = 3.072 \times 10^7 m_{♂}$$

From tabulated data in Bate, Mueller, and White [1], we have the answer to be 3.083×10^7. ♠

For the Earth-Moon system, the solution to the equations would be significantly influenced by the inclusion of the mass of the Moon. As the Earth is only about 81 times as massive as the Moon, the mass of the latter is not always negligible in calculations of this sort.

2.9 Canonical Units

So far, we have only seen units of length and time in examples. This is because the physics and formulae we have investigated are the same, whether measured in seconds or days, feet or furlongs. However, when it comes time to apply these results to systems, we need to have numbers to use, and those numbers are usually a count of some unit.

Some of the examples so far in this book have involved the Earth, and we have used the value $3.98601 \times 10^5\,\text{km}^3/\text{s}^2$ as the gravitational parameter. This in turn requires the use of kilometers for all length measurements and seconds for all time measurements, with the result that velocities are specified in kilometers per second. As a result, distances tend to be in the thousands and tens of thousands, velocity on the order of ten, and periods in the tens of thousands. For example, a circular low-Earth orbit might have radius $r = 6900\,\text{km}$, which gives $v_c = 7.600\text{km/s}$ and $P = 5704\,sec$. The orbital energy is $C = -28.88\,\text{km}^2/\text{s}^2$, and the angular momentum is $h = 52444\,\text{km}^2/\text{s}$.

These values become even more disparate when solar scales are involved. Distances are hundreds of millions of kilometers, velocities tens of kilometers per second, and periods in the millions of seconds (a mean solar year is over 31 million seconds).

The concept of *canonical units* unifies the various units, and also makes the results for various physical systems similar. The idea of a canonical unit system is

[9] Danby [6], appendix C.2.

that the actual units used for length and time are arbitrary, so that they may chosen to be convenient.

In the case of orbital systems, the physical system itself appears in the equations through the gravitational parameter. We can therefore normalize the units, by choosing the units *for a particular problem* such that the gravitational parameter is unity.

To demonstrate, we can go back to the equation of motion: $\ddot{\mathbf{r}} = -\mu \mathbf{r}/r^3$. $\mathbf{r}$ is measured in distance units, and the acceleration $\ddot{\mathbf{r}}$ in distance per time unit squared. Abbreviating, and writing the equation only with units, this gives

$$\frac{\mathrm{DU}}{\mathrm{TU}^2} = -\mu \frac{1}{\mathrm{DU}^2}$$

which implies that the units of μ are $\mathrm{DU}^3/\mathrm{TU}^2$, as in $\mathrm{km}^3/\mathrm{s}^2$. But we can define any set of units we like. Because mass does not appear in the calculations, we have only one unit to specify; the others will then be computed to be consistent.

The usual approach is to specify a distance unit. The distance unit for Solar System work is the *Astronomical Unit*, *AU* which is roughly the mean diameter of the orbit of the Earth about the Sun.[10] The time unit is then defined such that $\mu = 1$. To compute this, recall that

$$P = 2\pi \sqrt{\frac{a^3}{\mu}}$$

so that the period of a circular orbit with radius $a = 1\,\mathrm{DU}$, with $\mu = 1\,\mathrm{DU}^3/\mathrm{TU}^2$, is $P = 2\pi\,\mathrm{TU}$.

Example 2.4. Given that $1AU = 1.495979 \times 10^8\,\mathrm{km}$, compute the conversions from the canonical time unit and the canonical velocity unit $1\,\mathrm{DU}/\mathrm{TU}$ to kilometers, and kilometers per second, respectively.

Solution: From the preceeding discussion, we have

$$P = 2\pi\,\mathrm{TU} = 2\pi\sqrt{a^3/\mu_\odot}.$$

Since we are told that $a = 1AU = 1.495979 \times 10^8\,\mathrm{km}$, and we can look up the value of $\mu_\odot$ in Appendix C, we have

$$1\,\mathrm{TU} = \sqrt{\frac{(1.495979 \times 10^8\,\mathrm{km})^3}{1.32715 \times 10^{11}\,\mathrm{km}^3/\mathrm{s}^2}} = 5022600\,\mathrm{s}$$

One distance unit per time unit is then $1.495979 \times 10^8/5022600 = 29.785\,\mathrm{km/s}$. ♠

[10]Recall that using only the mass of the larger body in the gravitational parameter is an approximation, since it ignores the mass of the smaller body. This is close enough for most work, but for very precise measurments it is not. Therefore, the astronomical unit is more correctly defined as the distance AU such that $2\pi(AU)^{3/2}/(Gm_\odot)^{1/2}$ equals the period of the Earth. This is itself arbitrary, of course, and it woudl have been just as correct to define the distance unit as the measured average distance and adjust the time unit accordingly.

The actual tabulated values of the conversion factors (from [1], p. 429) are $1\,\mathrm{TU} = 5022675.7\,\mathrm{s}$ and $1\,\mathrm{DU/TU} = 29.784852\,\mathrm{km/s}$. These are calculated from a slightly different value of the astronomical unit.

Note: Various sources agree on these values to 5, 6, or more decimal places. Vallado [16], for example, lists

$$1\,\mathrm{TU} = 58.132440906\,\text{solar days} = 5022642.8943\,\mathrm{s}$$

and

$$1\,\mathrm{DU/TU} = 29.7846916749\,\mathrm{km/s}$$

We will seldom need such precision in this book. Because of the difficulty of choosing a particular authority, no definitive values for these constants are listed in this book. If needed, the values derived from the physical data in Appendix C will suffice. ♣

As mentioned, the distance unit for the Heliocentric system is the astronomical unit. The distance unit for the Geocentric system (in which the Earth is the central body) is the mean radius of the Earth (in this text, $r_\oplus = 6378.145\,\mathrm{km}$). The time unit is then computed as in the above example.

2.10 Problems

1. Classify the following orbits as circular, hyperbolic, *et cetera*. Assume canonical units.
   (a) $r = 2$, $v = 2$.
   (b) $r_p = 1.5$, $p = 3$.
   (c) $\mathbf{r} = 1.2\hat{k}$, $\mathbf{v} = -0.5\hat{k}$.
   (d) $C = -1/4$, $p = 2$.
   (e) $\mathbf{r} = 1.2\hat{\imath} + 0.5\hat{k}$, $\mathbf{v} = -0.25\hat{\imath} + 0.3\hat{\jmath} + 0.6\hat{k}$.
2. What is the eccentricity of an orbit for which $r_a = kr_p$, in terms of k? Check your result by showing that the cases $k = 1$ and $k \longrightarrow \infty$ give the appropriate values.
3. Find the ratio of the angular rate $d\nu/dt$ at periapsis to the rate at apoapsis, in terms of the orbital eccentricity.
4. What is the average value of r for an elliptic orbit, taking the true anomaly as the independent variable?
5. What velocity would the moon need to escape from the Earth, starting at its present orbital radius? Compute the answer both with and without including the mass of the moon in computing the gravitational parameter.

6. A vehicle is traveling near the Earth with velocity $v = 5000$ meters per second.
   (a) If the vehicle is in circular orbit, what is its altitude above the surface?
   (b) What is the closest it can be to the Earth if it is on an escape trajectory?
   (c) What is the orbital eccentricity if it is at its closest approach, and the distance to the center of the Earth is 24000 kilometers?

7. Take the orbit of the earth about the sun to be circular (a good first approximation), and consider a meteor approaching the earth on a parabolic orbit about the sun. What is the minimum and maximum speed of the meteor relative to the earth when it strikes?

8. Show that the velocity at the ends of the semiminor axis of an elliptic orbit is the same as the velocity of a satellite on a circular orbit that passes through that point.

9. Show that the flight path angle is 45° when the true anomaly $\nu = \pi/2$ for all parabolic orbits.

10. What is the value of r on an elliptic orbit such that the vehicle is moving most directly *away* from the central body (that is, $\mathbf{v}$ is most nearly aligned with $\mathbf{r}$)?

11. Halley's comet last passed perihelion on February 9, 1986. It has semimajor axis $a = 17.9654$ AU and eccentricity $e = 0.967298$.
    (a) Calculate the radius and velocity at perihelion and aphelion, in both canonical units and kilometers and seconds.
    (b) Calculate the period, and date of next perihelion.

12. Derive an expression for the period of an orbit in terms of the speed v_0 and radius r_0 at some specified point.

13. (Moulton [11]) Suppose a comet on a parabolic orbit with perihelion q strikes and combines with a body of equal mass, which is stationary before the collision. What is the eccentricity and perihelion distance of the combined mass?

14. (Wiesel [17]) Suppose that a spaceship with a solar sail is in motion about the sun. The sail is kept flat on to the sun, so that it produces an acceleration from radiation pressure of
$$\mathbf{a} = k\frac{\mathbf{r}}{r^3}$$
where k is a constant depending on sail reflectivity, vehicle mass, and other properties. Show that the equation of motion of the spaceship is
$$\ddot{\mathbf{r}} = -\left(\mu_\odot - k\right)\frac{\mathbf{r}}{r^3}.$$

What types of consic sections are possible if $k < \mu_\odot$? When they are equal? When $k > \mu_\odot$?

15. Assume a comet is on a parabolic trajectory, which lies in the plane of the Earth's orbit about the sun. What perihelion maximizes the amount of time that the comet spends within the orbit of the Earth? Assume the orbit of the Earth is circular.

Chapter 3

Position in Orbit and Time in Transit

In the preceding chapter, we have derived the shape and size of orbits. This allows us to predict where a spacecraft or celestial body will travel, and conversely, where it has been, at least to the accuracy of the two-body assumptions we have made. It remains to express the position on the orbit as a function of time, so that we can answer the question not only of *where* the craft is going, but also *when* it will get there.

We will first consider these questions for the elliptic orbit, for which we can find a graphical approach to the solution. After that, we will explore the hyperbolic orbit, and then the parabolic as the limiting case of the other two. We will find expressions for the position as a function of time, and for the time of transit between two points on a particular orbit. In the final section of the chapter, we will look at some numerical methods for finding the solutions to the equations we derive. This will complete the necessary machinery for the solution of the mission problems in later chapters.

3.1 Kepler's Equation

We will begin by considering the elliptic orbit. The approach we take is due to Kepler, and is geometric in nature. It is based on Kepler's second law, that the rate at which area is swept along an orbit is constant. This means that if the area is known, the time it took to sweep the area is known. Unfortunately, the area under an ellipse is not a simple thing to calculate. Therefore, Kepler derived a solution in terms of a superscribed circle, as in Figure 3.1a.

First note that we can relate areas in ellipses to areas in circles. We do this by noting that in Figure 3.1a, a vertical line such as PR has the property that

$$\frac{\overline{PQ}}{\overline{PR}} = \frac{b}{a}$$

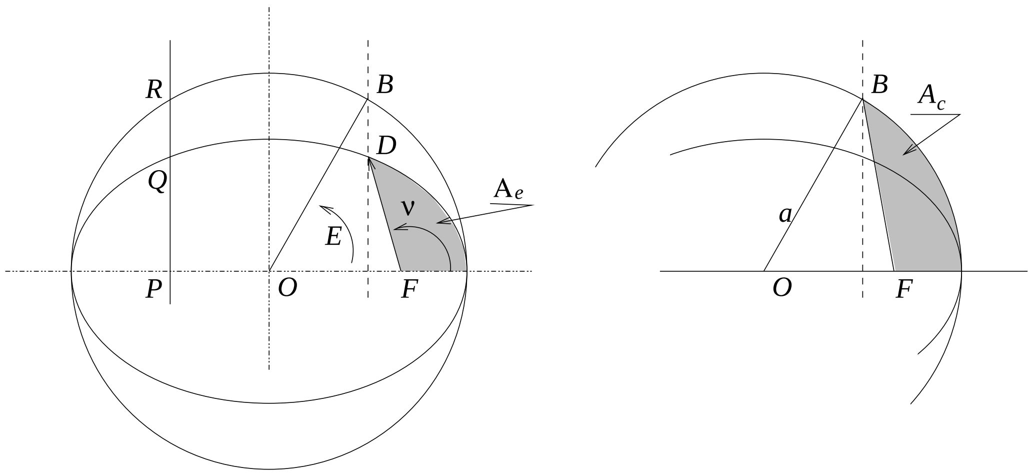


Figure 3.1: Deriving Kepler's equation.

The shaded area A_c in Figure 3.1b is thus proportional to A_e in Figure 3.1a. For a circle, however, the area is much easier to calculate.

Kepler defined the angle E as the *eccentric anomaly* (as opposed to the true anomaly ν). The area enclosed by the circle, the major axis, and the radius OB of the circle is simply $a^2E/2$ (E in radians). Then we subtract the area of the triangle, and we have

$$\begin{aligned} A_c &= \frac{1}{2}a^2E - \frac{1}{2}a^2 e \sin E \\ &= \frac{1}{2}a^2(E - e\sin E) \end{aligned}$$

Then the area we wanted in the first place is just

$$A_e = \frac{bA_c}{a} = \frac{ab}{2}(E - e\sin E)$$

Now recall Kepler's 2nd and 3rd laws. The third law gives us

$$T = 2\pi\sqrt{a^3/\mu}$$

and we know that the total area of an ellipse is πab. The 2nd law gave us that dA/dt is constant. This means that it must equal the total area divided by the period, so

$$\frac{dA}{dt} = \frac{\pi ab}{2\pi\sqrt{a^3/\mu}} = \frac{ab}{2}\sqrt{\frac{\mu}{a^3}}$$

Now, A_e must equal $\dot{A}(t - t_p)$, where t_p is the time at periapsis. Plugging our last result into this gives

$$A_e = \frac{ab}{2}(E - e\sin E) = \frac{ab}{2}\sqrt{\frac{\mu}{a^3}}(t - t_p) \tag{3.1}$$

We now make two further definitions. The first is the *mean motion* n (sometimes called the *mean angular velocity*), which is simply the average value of the angular velocity. This is the total angle swept in a single orbit, divided by the period of the orbit, and is given by

$$n = 2\pi/T = \sqrt{\mu/a^3}$$

The second is the *mean anomaly* M

$$M = n(t - t_p)$$

Using these definitions, 3.1 becomes

$$M = E - e\sin E$$

This result is known as *Kepler's Equation.* In terms directly of time since periapsis passage, this can be written

$$t - t_p = \sqrt{a^3/\mu}\,(E - e\sin E) \tag{3.2}$$

and we will see this many times in future chapters.

Assuming we can solve Kepler's equation, we can use the result to find ν, which is what we started out looking for. Going back to Figure 3.1a, we can see that

$$a\cos E - r\cos\nu = ae$$

Recalling the orbit equation, we have

$$\begin{aligned} \cos E \;&=\; \frac{p\cos\nu}{a(1+e\cos\nu)} + e \;=\; \frac{a(1-e^2)\cos\nu}{a(1+e\cos\nu)} + \frac{e(1+e\cos\nu)}{1+e\cos\nu} \\ &\Longrightarrow \cos E = \frac{e+\cos\nu}{1+e\cos\nu} \end{aligned} \tag{3.3}$$

Now, the altitude of the inscribed triangle is related to its intersection with the ellipse as

$$(b/a)a\sin E = r\sin\nu$$

We again plug in for r and rearrange, arriving at

$$\sin E = \frac{\sqrt{1-e^2}\sin\nu}{1+e\cos\nu} \tag{3.4}$$

We rearrange the expression (3.3) for $\cos E$ into one for $\cos\nu$ instead:

$$\cos\nu = \frac{e - \cos E}{e\cos E - 1}$$

Now substituting this into equation (3.4) for and manipulating, we finally get

$$\sin\nu = \frac{\sqrt{1-e^2}\sin E}{1-e\cos E}$$

An alternate formulation is

$$\tan\frac{\nu}{2} = \sqrt{\frac{1+e}{1-e}}\tan\frac{E}{2} \tag{3.5}$$

which is very convenient, as $\nu/2$ and $E/2$ are always in the same quadrant.

This expression is valid only for elliptic orbits. The expression for hyperbolic orbits is similar, and we will derive it later. First, however, we will look at the time of transit on an elliptic orbit.

3.2 Lambert's Problem

Having an expression for the position along an orbit as a function of time is very much like having solved the time in transit problem. We can consider the position problem as the time in transit from one point to another, where the initial point is the periapsis. However, this would require that we know the eccentricity of the orbit. As it turns out, this is not necessary.

The problem of transit time between two points on an elliptic orbit is often referred to as *Lambert's Problem*, after Johann Heinrich Lambert (1728 – 1779). Lambert explored the problem in depth, and in 1761 postulated what has become known as *Lambert's Theorem*, which can be stated:

> The time required to travel between two points on an elliptic arc depends only on the semimajor axis of the ellipse, the straight-line distance between the two points, and the sum of the distances of the points from the focus of the ellipse.

This is sketched in Figure 3.2. The values needed for Lambert's problem are the distance d, the semimajor axis, and the sum $r_1 + r_2$. The first analytical proof of this theorem was not in fact due to Lambert, but was provided by Lagrange in 1778.

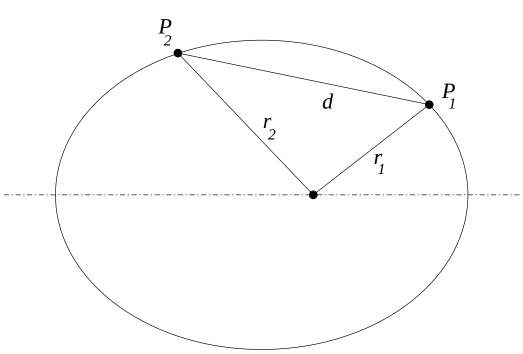


Figure 3.2: The elliptic transit problem.

In this section, we will derive Lambert's result using the geometric properties of the ellipse. This will result in an equation that can be used to solve for the transfer time given the semimajor axis of the transfer orbit. The same equation can be used with minor modifications for parabolic and hyperbolic orbits as well.

Lambert's problem concerns the time it takes to travel from one point to another on an elliptic orbit. This problem is clearly similar to that of position as a function of time, in which we

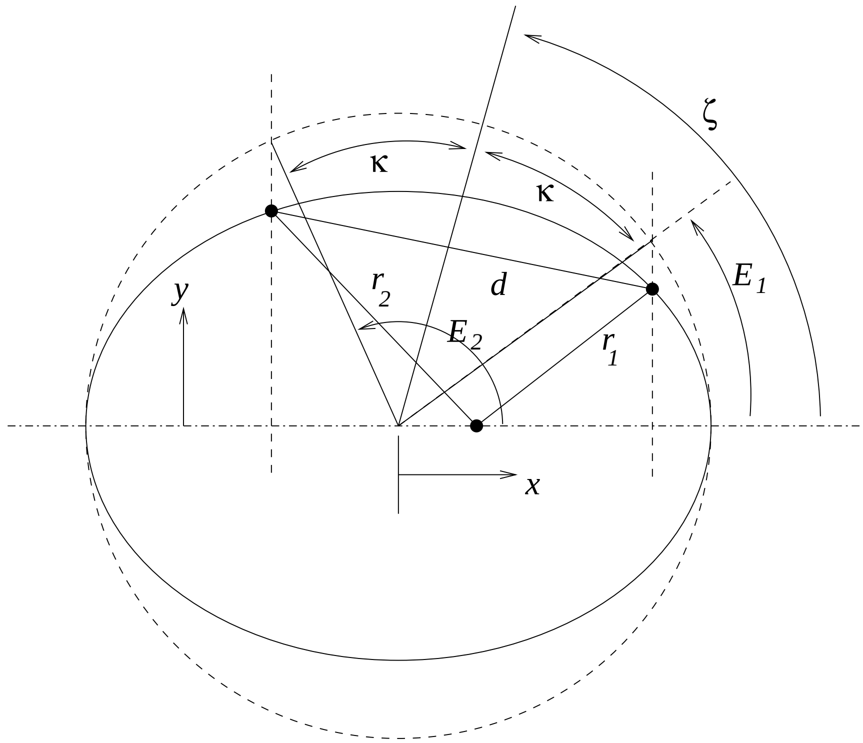


Figure 3.3: Solving Lambert's problem.

found it convenient to work with eccentric anomaly. Thus, we set up the problem as in Figure 3.3. We will assume that E_2 is greater than E_1. This is always reasonable, since if it is not the case, we can solve the identical problem of flight from P_2 to P_1 on an identical orbit in the opposite direction. It will turn out to be useful to define the angles

$$\begin{aligned}\zeta &= \frac{1}{2}(E_2 + E_1) \\ \kappa &= \frac{1}{2}(E_2 - E_1)\end{aligned}$$

and perform the analysis using these rather than the eccentric anomalies themselves.

First consider the distance between the points. We can see from Figure 3.3 that the x and y coordinates of any point on the ellipse are given by

$$x = a\cos E \qquad \text{and} \qquad y = b\sin E \tag{3.6}$$

The distance is given by $d^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$, which can be written (using the definitions above) as

$$d^2 = a^2\left[\cos(\zeta+\kappa) - \cos(\zeta-\kappa)\right]^2 + b^2\left[\sin(\zeta+\kappa) - \sin(\zeta-\kappa)\right]^2 \tag{3.7}$$

Recalling $b^2 = a^2(1 - e^2)$ and simplifying the trigonometric terms leads to

$$\begin{aligned} d^2 &= a^2\left[(\cos\zeta\cos\kappa - \sin\zeta\sin\kappa) - (\cos\zeta\cos\kappa + \sin\zeta\sin\kappa)\right]^2 \\ &\quad + a^2(1 - e^2)\left[(\sin\zeta\cos\kappa + \cos\zeta\sin\kappa) - (\sin\zeta\cos\kappa - \cos\zeta\sin\kappa)\right]^2 \\ &= 4a^2\sin^2\zeta\sin^2\kappa + 4a^2(1 - e^2)\cos^2\zeta\sin^2\kappa \\ &= 4a^2\sin^2\kappa(1 - e^2\cos^2\zeta) \end{aligned} \tag{3.8}$$

Because $e < 1$, the product $e\cos\zeta < 1$, so we invent an angle ξ and define

$$\cos\xi = e\cos\zeta \tag{3.9}$$

Using this, we can rewrite 3.8 as

$$\begin{aligned} d^2 &= 4a^2\sin^2\kappa(1 - \cos^2\xi) \\ &= 4a^2\sin^2\kappa\sin^2\xi \\ &\Longrightarrow \quad d = 2a\sin\kappa\sin\xi \end{aligned} \tag{3.10}$$

Using the fact that

$$r = a(1 - e\cos E) \tag{3.11}$$

and 3.9, the sum of the radii can be expressed as

$$\begin{aligned} r_1 + r_2 &= 2a - ae\cos(\zeta - \kappa) - ae\cos(\zeta + \kappa) \\ &= 2a(1 - e\cos\zeta\cos\kappa) \\ &= 2a(1 - \cos\kappa\cos\xi) \end{aligned} \tag{3.12}$$

Now add this expression to the result (3.10) for the distance.

$$r_1 + r_2 + d = 2a(1 + \sin\kappa\sin\xi - \cos\kappa\cos\xi) \tag{3.13}$$

Recognizing the right hand side as having the form of a trigonometric angle-sum identity, we define

$$\alpha = \xi + \kappa$$

which allows 3.13 to be written as

$$\begin{aligned} r_1 + r_2 + d &= 2a(1 - \cos\alpha) \\ &= 4a\sin^2(\alpha/2) \end{aligned} \tag{3.14}$$

Similarly, defining

$$\beta = \xi - \kappa$$

allows the creation of the identity

$$r_1 + r_2 - d = 2a(1 - \cos\beta) = 4a\sin^2(\beta/2) \tag{3.15}$$

Equations (3.14) and (3.15) allow us to use the known values of a, r_1, r_2, and d to solve for α and β, and thereby for the eccentric anomalies E_1 and E_2, if we also know the eccentricity. This would be all that we need to find the time of travel between the two points using Kepler's equation. The transit time can be written

$$\Delta t = t_2 - t_1 = \sqrt{a^3/\mu}\,[(E_2 - e\sin E_2) - (E_1 - e\sin E_1)]$$

Writing the eccentric anomalies in terms of ζ and κ produces

$$\begin{aligned}
\sqrt{\mu}\Delta t &= a^{3/2}\,[2\kappa + e\sin(\zeta-\kappa) - e\sin(\zeta+\kappa)] \\
&= a^{3/2}\,[2\kappa - 2e\sin\kappa\cos\zeta] \\
&= a^{3/2}\,[\alpha - \beta - 2\sin\kappa\cos\xi] \\
&= a^{3/2}\,[\alpha - \beta - (\sin\alpha - \sin\beta)] \qquad (3.16)
\end{aligned}$$

Here, the eccentricity turns out not to be necessary, as we have managed to write the time of transit in terms only of a, α, and β. Grouping the significant results, we have

$$\Delta t = \sqrt{a^3/\mu}\,[(\alpha - \sin\alpha) - (\beta - \sin\beta)] \qquad (3.17)$$

$$\alpha = 2\sin^{-1}\sqrt{\frac{r_1 + r_2 + d}{4a}} \qquad (3.18)$$

$$\beta = 2\sin^{-1}\sqrt{\frac{r_1 + r_2 - d}{4a}} \qquad (3.19)$$

(3.17) is often called *Lambert's Equation*. For this set of equations to be valid, the transfer must match the assumptions used to derive them. Recall that we assumed a transfer from a point nearer perigee to one farther away. Also, we have tacitly assumed that the satellite does not pass perigee during the transfer. These assumptions will be taken into account when we discuss the implementation of Lambert's work to solving problems in the last section of the chapter.

3.3 Hyperbolic Trajectories

In obtaining our results for time of flight and time in transit for elliptic orbits, it was helpful to be able to use geometric intuition. This is not as easy in the case of hyperbolic trajectories, as there is no superscribed figure with simple, familiar properties to use. Instead, we will use more algebraic approaches, guided by the similarity between these problems and the elliptic problem.

3.3.1 An Analogue of Kepler's Equation

For hyperbolic orbits, the semimajor axis is negative. It will turn out to be convenient in the derivations of this section to replace a with its absolute value, which will be denoted

$$\bar{a} \triangleq -a$$

Then we begin with *vis-viva* equation

$$v = \sqrt{\frac{2\mu}{r} - \frac{\mu}{a}} = \sqrt{\frac{2\mu}{r} + \frac{\mu}{\bar{a}}}$$

and substitute the components of the velocity and square the equation to get

$$v^2 = \dot{r}^2 + r^2\dot{\nu}^2 = \frac{2\mu}{r} + \frac{\mu}{\bar{a}} \tag{3.20}$$

Recall from the derivation of Kepler's second law the result

$$2\dot{A} = r^2\dot{\nu} = h$$

and rewrite this to get

$$\begin{aligned} r^2\dot{\nu} &= \sqrt{\mu p} = \sqrt{\mu\bar{a}(e^2-1)} \\ &\implies \dot{\nu}^2 = \frac{\mu\bar{a}(e^2-1)}{r^4} \end{aligned} \tag{3.21}$$

Using eqn. (3.21) to eliminate $\dot{\nu}$ from eqn. (3.20) gives

$$\begin{aligned} \dot{r}^2 &= \frac{2\mu}{r} + \frac{\mu}{\bar{a}} - \frac{\mu\bar{a}(e^2-1)}{r^2} \\ &\implies (\bar{a}r^2/\mu)\dot{r}^2 = 2\bar{a}r + r^2 - \bar{a}^2e^2 + \bar{a}^2 = (r+\bar{a})^2 - \bar{a}^2e^2 \\ &\implies \sqrt{\frac{\bar{a}}{\mu}}\frac{rdr}{dt} = \sqrt{(r+\bar{a})^2 - \bar{a}^2e^2} \end{aligned} \tag{3.22}$$

We can define a variable analogous to the mean motion for this case. Recalling that $n = \sqrt{\mu/a^3}$, define

$$\bar{n} = \sqrt{\mu/\bar{a}^3}$$

Then from eqn. (3.22) we can write

$$\begin{aligned} \frac{\sqrt{\bar{a}^3/\mu}}{\bar{a}}\frac{rdr}{dt} &= \sqrt{(r+\bar{a})^2 - \bar{a}^2e^2} \\ \implies \bar{a}\bar{n}dt &= \frac{rdr}{\sqrt{(r+\bar{a})^2 - \bar{a}^2e^2}} \end{aligned} \tag{3.23}$$

This equation can be integrated at least once in terms of hyperbolic functions. Knowing this, we are moved to introduce a value similar to the eccentric anomaly. Denoting this value F, we define it through the relation

$$r = \bar{a}(e\cosh F - 1) \tag{3.24}$$

(note the similarity to eqn. (3.6)), and refer to it as the *hyperbolic eccentric anomaly.* From eqn. (3.24), we see that

$$\cosh F = \frac{\bar{a} + r}{\bar{a}e} \qquad \text{and} \qquad dr = \bar{a}e \sinh F dF$$

which allows eqn. (3.23) to be written (after quite a bit of manipulation) as

$$\bar{n}dt = (e \cosh F - 1)dF$$

This relation is simple to integrate, and results in

$$e \sinh F - F = \bar{n}(t - t_p) \tag{3.25}$$

which is the hyperbolic equivalent of Kepler's equation.

Beginning with the relation

$$r = a(1 - e \cosh F) = \frac{a(1 - e^2)}{1 + e \cos \nu}$$

it is straightforward to derive equations such as

$$\cos \nu = \frac{\cosh F - e}{1 - e \cosh F} \tag{3.26}$$

relating the true anomaly to the hyperbolic eccentric anomaly. Further such relations are left as an exercise for the reader.

3.3.2 Time in Transit

Deriving Lambert's Equation for elliptic orbits required a goodly amount of geometric intuition, as well as some fairly intricate algebra. We will use the insight gained in that derivation to help in finding a solution to the problem in the case of hyperbolic orbits.

One very useful bit of information is that there exist analogues of eqn. (3.6) and eqn. (3.11) for the hyperbola. We have already defined the hyperbolic eccentric anomaly F; it turns out that the hyperbola obeys the relations[1]

$$x = \bar{a} \cosh F \qquad \text{and} \qquad y = \bar{a}\sqrt{e^2 - 1} \sinh F \tag{3.27}$$

It is easy enough to show that eqn. (3.24) is satisfied under these relations.

Note that F is *not* an angle in the Euclidean sense, as is the eccentric anomaly E. F does not range from 0 to 2π as does an angle in standard geometry. Rather, this angle is defined on a hyperbola, and runs from $-\infty$ to $+\infty$.

[1] When $e = \sqrt{2}$, the result is an *equilateral hyperbola.* The results of this section can be derived through geometrical arguments using the equilateral hyperbola in the place of the circumscribed circle used to derive Kepler's equation.

Once this relationship is defined, steps nearly identical to those performed to derive Lambert's equation can be followed to get an analogous result for the hyperbola. The resulting equation is

$$\Delta t = \sqrt{\bar{a}^3/\mu}\,[(\sinh\alpha - \alpha) - (\sinh\beta - \beta)] \tag{3.28}$$

where in this case the angles α and β are given by

$$\alpha = 2\sinh^{-1}\sqrt{\frac{r_1 + r_2 + d}{4\bar{a}}} \tag{3.29}$$

$$\beta = 2\sinh^{-1}\sqrt{\frac{r_1 + r_2 - d}{4\bar{a}}} \tag{3.30}$$

Recall that we are using hyperbolic angles and functions. Therefore, α and β are not restricted to the range $[0, 2\pi]$.

3.4 Parabolic Trajectories

The parabola is in many ways the easiest trajectory for which to derive a relationship for position as a function of time. Beginning with the orbit equation

$$r = \frac{p}{1 + \cos\nu}$$

we use the facts that $p = h^2/\mu$ and $1 + \cos\nu = 2\cos^2(\nu/2)$ to write

$$r = \frac{h^2}{2\mu}\frac{1}{\cos^2(\nu/2)}$$

We recall from the derivation of Kepler's Laws that

$$\begin{aligned} r^2\frac{d\nu}{dt} &= h \\ &\Rightarrow dt = \frac{r^2}{h}d\nu \end{aligned}$$

so that we now have

$$dt = \frac{h^3}{4\mu^2}\cos^4\left(\frac{\nu}{2}\right)\,d\nu$$

which in turn implies

$$t - t_p = \frac{h^3}{4\mu^2}\int_0^\nu \sec^4\left(\frac{\nu}{2}\right)\,d\nu$$

This equation can be integrated to get the relation

$$t - t_p = \frac{h^3}{2\mu^2}\left(\tan\frac{\nu}{2} + \frac{1}{3}\tan^3\frac{\nu}{2}\right) \tag{3.31}$$

This result is known as *Barker's Equation*. Barker did not establish the relationship, but provided extensive tables for its solution during the eighteenth century.

One of the great benefits of Lambert's work is that the eccentricity is not needed to solve for time of flight. This is not a great contribution in the case of parabolic trajectories, as the eccentricity is known for this case. Theoretically, one could iterate on Barker's equation to solve for time of flight. In this case, the result in the style of Lambert produces a simple equation, instead.

Beginning with the elliptic time of flight equations, we take the limit as $a \to \infty$. Recalling (3.18), we can see that the sine of α and thus α itself becomes small. This holds true also for β. We can therefore replace the sines of the angles with the Taylor expansion about zero to get

$$\Delta t = \sqrt{a^3/\mu}[(\alpha^3/6 - \cdots) - (\beta^3/6 - \cdots)]$$

When taking the limit, we will see that all higher order terms go to zero, and after substitution arrive at the result

$$\Delta t = \frac{(r_1 + r_2 + d)^{3/2} - (r_1 + r_2 - d)^{3/2}}{6\sqrt{\mu}} \tag{3.32}$$

This equation, having been obtained from the expression derived for elliptic orbits, contains all of the assumptions inherent in the earlier result, and more. As a result, (3.32) is valid only when the angle between the two radial vectors is less than π; when the angle is greater than π, the minus sign is converted to a plus and the relationship is

$$\Delta t = \frac{(r_1 + r_2 + d)^{3/2} + (r_1 + r_2 - d)^{3/2}}{6\sqrt{\mu}} \tag{3.33}$$

This and other issues involved in solving time of flight problems are addressed in the following sections.

3.5 Implementation Considerations

Although the equations derived in this section are framed in the context of finding time of flight, it is also very common to see the inverse problem. That is, given two (or more) observations, find the semimajor axis and eccentricity of the orbit. This is a problem of orbit determination.

In general, we would expect the information we are given to consist of the distances to the object, the times of the observations, and the angle between the radial vectors at the two times. Knowing the angle, we can construct the chord length, giving us most of the information necessary for the time of flight solution. The final part of the problem is to iterate on the value of the semimajor axis until a value is found that gives a time of flight that matches the known actual time.

Using the results of the previous sections to solve for Δt is not always straightforward. The equations include trancendental functions which must be solved numerically, and the problem may not fit the assumptions used in deriving the equations. Further, the results (except when using universal variables) give different equations for elliptic, hyperbolic, and parabolic orbits, which necessitates some way to know which set of equations to use. Even when the type of conic section is known, there may be difficulties.

In solving position and time of flight problems, we will usually find it necessary to use some numerical methods. The class of techniques we will need are those involved in finding the zeroes of functions. As this will be a common feature of the problems ahead, we will cover this topic next.

3.5.1 Finding the Zeroes of Functions

The general class of problems we consider here are stated: Given a known function $f(x)$, find one or more values of x such that $f(x) = 0$. There are many such techniques, but we will consider only two.

Note that in numerical methods, we do not find actual values of the zeroes. Instead, we find values that sufficiently closely approximate the them. The accuracy required is generally decided based on the needs of the particular problem.

Bisection

The most basic of all zero-finding techniques is *bisection*. This method usually assumes that there are two known values $\underline{x}$ and $\overline{x}$ such that

$$f(\underline{x}) \cdot f(\overline{x}) < 0 \tag{3.34}$$

We assume that $f(x)$ is continuous on the interval $[\underline{x}, \overline{x}]$, so that this condition guarantees that $f(\tilde{x}) = 0$ for some $\tilde{x} \in [\underline{x}, \overline{x}]$. We will further assume that there is only one such value, though this is not actually necessary.

The technique is simplicity itself. Written in algorithmic format, it might be described

$$
\begin{array}{rl}
\text{Given :} & \underline{x},\ \overline{x},\ \varepsilon > 0 \\
(\star) & \tilde{x} = (\underline{x} + \overline{x})/2 \\
 & \text{if } |f(\tilde{x})| < \varepsilon \ \text{ exit} \\
 & \text{if } f(\tilde{x}) \cdot f(\underline{x}) < 0 \text{ then} \\
 & \quad \overline{x} \leftarrow \tilde{x} \\
 & \text{else} \\
 & \quad \underline{x} \leftarrow \tilde{x} \\
 & \text{go to } (\star)
\end{array}
$$

This is read something like: "Given a lower bound, an upper bound, and a tolerance ε, set $\tilde{x}$ to the midpoint of the interval. If the value of the function is sufficiently small, call $\tilde{x}$ a zero and exit. If not, check to see if the zero is between the lower bound and $\tilde{x}$. If it is, set the upper bound to $\tilde{x}$. If it is not, set the lower bound to $\tilde{x}$. Then go back to the top and do it all over again."

It is easy to see that the length of the interval will be halved at each step, giving rise to the name of the technique. The value of ε is an error tolerance; when the absolute value of the function is less than ε we take it to be effectively zero.

The version of the technique given here is the most basic. It is common to add another tolerance, this one on the length of the interval. In this way, it is possible to control the accuracy with which $\tilde{x}$ is found, which is often more important than how close the function is to zero. It also puts a bound on the number of iterations that will be needed to find a value of $\tilde{x}$ that is "close enough" to the actual zero of the function.

Newton-Raphson Iteration

The bisection technique is the simplest and in many ways the most robust of the zero-finding algorithms. All it requires is an interval, a tolerance, and some time. It is also the slowest of the techniques. It is almost never used in practice.

The most popular technique in this class is known as *Newton-Raphson iteration.* This is based on the Taylor expansion of function under consideration about the current point x:

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f'' + \cdots$$

Since we are looking for the value $\tilde{x}$ for which $f(\tilde{x}) = 0$, we use the first two terms of the expansion to approximate

$$f(\tilde{x}) = f(x+h) \approx f(x) + hf'(x) = 0$$

We can then easily solve for h as

$$h = -\frac{f(x)}{f'(x)}$$

The algorithm is usually written something like

$$\begin{array}{rl}
\text{Given :} & f(\cdot),\ f'(\cdot),\ x_0,\ \epsilon > 0 \\
 & x \leftarrow x_0 \\
(\star) & \text{If } |f(x)| < \varepsilon \text{ exit} \\
 & h \leftarrow f(x)/f'(x) \\
 & x \leftarrow x - h \\
 & \text{go to } (\star)
\end{array}$$

Generally, a count of iterations is kept and the procedure aborted if the number gets very large.

Again, there are many enhancements to the basic algorithm presented. It is standard to include a value of η as a lower bound on the magnitude of h; when the correction to the current estimate of the zero is less than η, the value is "close enough".

Example 3.1. Find $\sqrt{2}$ using Newton-Raphson iteration.

Solution: Finding the square root of a number y is equivalent to finding x such that $y - x^2 = 0$. Therefore we set up our equations with $y = 2$ as

$$\begin{aligned} f(x) &= 2 - x^2 \\ f'(x) &= -2x \end{aligned}$$

Taking $x_0 = 1$ as our starting guess, we have the sequence of approximations

x	$f(x)$	$f'(x)$	h
1	1	-2	-0.5
1.5	-0.25	-3	$0.08333\cdots$
$1.41666\cdots$	$-0.006944\cdots$	$-2.83333\cdots$	0.002451
1.414216	-6.0×10^-6	–	–

Here we have chosen to stop looking when the value of the function is zero to 6 digits.

Note the speed with which the algorithm converges to the solution. Note also, however, that had the starting value been $x_0 = 0$, the first iteration would have resulted in $f' = 0$, causing the algorithm to fail. ♠

Newton-Raphson iteration is almost always much, much faster than bisection, when it works. It requires more work, in that the derivative of the function must be evaluated as well as the function itself, but this is usually a small price to pay for the much smaller number of total iterations. The drawback to the technique is that, for general functions there is no guarantee that the derivative will not be zero on the interval, in which case the technique will fail. There is also the possibility that if x_0 is poorly chosen, the technique will take some time to come near the desired zero, and may in fact find the wrong zero altogether.

In our cases, the functions we will consider will be well-behaved, and we will not need to worry. We will use the Newton-Raphson technique rather than bisection, with some thought given to finding appropriate initial guesses at the zero positions.

3.5.2 Solving Kepler's Equation

The only real difficulty in applying the Newton-Raphson technique to Kepler's equation is the necessity to find a starting point. This was particularly important in the years before computers, of course, when all computations had to be performed by hand. As a result, there is an extensive literature devoted to the subject. We will make only a couple of elementary observations here.

We first note that, as $e \to 0$, the ellipse becomes a circle. On a circular orbit, $E = M$, so we might suspect that for small eccentricity, E will be close to M. This is a fair assumption in many cases.

Example 3.2. The eccentricity of Mars' orbit is $e = 0.0934$. Find ν 200 days[2] after perihelion.

Solution: First find the mean anomaly M. For simplicity, we will use tabulated values of Mars' orbital period:

$$T_{♂} = 1\text{y}\,321.73\text{d} = 687.0\text{d} \quad \Longrightarrow \quad M = 2\pi\frac{200.0}{687.0} = 1.8292$$

Writing Kepler's equation in a form suitable for Newton-Raphson gives

$$M = E - e\sin E \Longrightarrow f(E) = E - e\sin E - M$$

The derivative $f'(E) = 1 - e\cos E$

Since e is small, we take as our first approximation $E = M$. Then the Newton-Raphson algorithm produces

E	$f(E)$	$f'(E)$	h
1.8292	−0.09030	1.0239	−0.088195
1.9174	0.00035	1.0317	0.000338
1.9170	5×10^{-9}	⋯	⋯

After only a single iteration, we have the value of E to four significant digits; after a second, the accuracy has increased to eight. We then solve for ν using eqn. (3.5):

$$\tan\frac{\nu}{2} = 1.5637 \quad \Longrightarrow \quad \nu = 2.0036$$ ♠

Setting $E_0 = M$ actually works quite well even for eccentricities in the range of $e = 0.5$. However, for values of e approaching unity, a better approximation can be made when it is realized that $M \le E \le M + e$. Convergence can be improved by choosing

$$E_0 = M + e/2$$

halfway between the bounds.

When solving equations on hyperbolic orbits, the same iteration techniques are used. However, the initial approximation may be harder to find. For e only slightly larger than unity, we can expand the hyperbolic sine term in eqn. (3.25) to get

$$M \approx e(F + F^3/6) - F \quad \Longrightarrow \quad F \approx \sqrt[3]{M/6}$$

[2]The use of the term *day* is confusing. In this case, because we have not specified sidereal days, we refer to the *mean solar day*. Confusingly enough, many tables list sidereal periods in mean solar units. Thus the sidereal period of the Earth's orbit is often listed as 365.256d.

As F becomes large, this approximation becomes less valid. Note that an upper bound on F is less easy to define in this case as well, as $\sinh F$ is not bounded.

When e is not small, we can begin with the same expansion, but retain e in the simplification:

$$e(F + F^3/6) - F - M \approx 0 \quad \Longrightarrow \quad F_0^3 + \frac{6(e-1)}{e}F_0 - \frac{6M}{e} = 0 \tag{3.35}$$

Solving this cubic will result in a better choice for F_0.

When F is large, the cubic approximation to $\sinh F$ becomes poor. Considering the exponential expression of $\sinh F$, we have[3]

$$\sinh F = \frac{\exp(F) - \exp(-F)}{2}$$

which, for large F, approaches $\sinh F = \exp(F)/2$. Thus, when F is large, we might use

$$F_0 = \ln(2M/e) \tag{3.36}$$

Which equation to use is hard to know *a priori*, since we do not know in advance whether F is large. To choose, we pick some test value of F. As in Roy [13], we try $F = 5/2$ and the expansion eqn. (3.35) becomes

$$M = e\left(\frac{5}{2} + \frac{125}{48}\right) - \frac{5}{2} \approx 5e - 5/2 \tag{3.37}$$

If M is larger than this approximation, then $F > 5/2$ and we use eqn. (3.36) to generate our first guess F_0. If M is less, then F is also less and we can use eqn. (3.35).

Example 3.3. Suppose we launch a space probe and accelerate it with a fly-by of Venus. If the probe reaches perihelion with a radius of 80×10^6km and speed $60\frac{\text{km}}{\text{s}}$ on January 1, 2000, on what day will it reach the orbit of Saturn?

Solution: This problem does not require iterative solution, because the angles are known. All that is required is to first compute the eccentricity:

$$C = v^2/2 - \mu_\odot/r = 141.25$$

[3]Since most calculators do not have hyperbolic function keys, it is convenient to recall the expressions

$$\sinh F = \frac{\exp(F) - \exp(-F)}{2} \qquad \cosh F = \frac{\exp(F) + \exp(-F)}{2}$$

$$\sinh^{-1} x = \ln(x + \sqrt{x^2+1}) \qquad \cosh^{-1} x = \ln(x + \sqrt{x^2-1})$$

$$\tanh^{-1} x = \frac{1}{2}\ln\left(\frac{1+x}{1-x}\right)$$

$$\begin{aligned} &\implies \quad a = -4.697 \times 10^8 \text{km} \\ h \quad = \quad & r_p v_p = 4.8 \times 10^9 \\ &\implies \quad e = \sqrt{1 + \frac{2Ch^2}{\mu^2}} = 1.1703 \end{aligned}$$

We know the radius of the orbit of Saturn is $a_{♄} = 9.5388 AU = 1.427 \times 10^9$km. We can then use the relation

$$r = a(1 - e \cosh F) \quad \implies \quad \cosh F = \frac{a - r}{ae} = 3.450$$

from which we find $F = 1.910$. We use this in (3.25) to find the time since perihelion.

$$t - t_p = \frac{e \sinh F - F}{\bar{n}} = 54.63 \times 10^6 \text{s} = 632.26 \text{ days}$$

Since the year 2000 is a leap year, it has 366 days and the probe will cross the orbit of Saturn on the 266th day of 2001. This turns out to be September 23. ♠

3.5.3 Time of Flight Problems

Solving time of flight problems is slightly more complicated than solving position problems, because we often do not know *a priori* whether the orbit is elliptic or hyperbolic. There are simple tests for this, however. It can be shown that the parabolic case is the limiting case for time of transfer; all other things being equal, the parabolic time of flight is the lower bound for the elliptic time of flight, and the upper bound for the hyperbolic.

Example 3.4. A comet is sighted at a distance of $r_1 = 6.336 \times 10^8$km from the Sun. 110 days later, it is sighted again, having traversed an angle of 20.9°, at a distance of $r_2 = 1.886 \times 10^8$km. Find the semimajor axis of the comet's trajectory.

Solution: We have to begin by figuring out what set of equations to use. We do this by trying out the parabolic elapsed time equation first.

Find the chord length: $d^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos(21°) \quad \implies \quad d = 4.623 \times 10^8$

The parabolic time of flight is given by eqn. (3.32) (note that the angle traversed is less than π) as

$$\Delta t = 1.794 \times 10^7 \text{s} \approx 208 \, \text{days}$$

Since the observed time of 110 days is less than this value, the trajectory must be hyperbolic.

We now turn to finding the semimajor axis, which will give us all we need to solve the problem. We have discussed no method by which a first guess is to be made. Here, we will simply try $\bar{a} = r_1 + r_2$, which results in

$$\Delta t(\bar{a} = 8.222 \times 10^8) = 1.589 \times 10^7 \text{s} \approx 184 \, \text{days}$$

This is far too high. Since the longest time of flight is given by $a \longrightarrow -\infty$, we suspect that moving a towards zero might shorten the time of flight. Since we need to shorten it considerably, let's try $\bar{a} = 3.00 \times 10^8$. This gives

$$\Delta t(\bar{a} = 3.00 \times 10^8) = 1.367 \times 10^7 \mathrm{s} \approx 158 \,\mathrm{days}$$

Since a reduction of over half in the semimajor axis took us less than halfway to the goal, we might try reducing it by a factor of 4 this time:

$$\Delta t(\bar{a} = 7.50 \times 10^7) = 9.221 \times 10^6 \mathrm{s} \approx 107 \,\mathrm{days}$$

A couple more guesses and we arrive at $a = -8.0 \times 10^7 \mathrm{km}$. ♠

Note that it is certainly possible to set up time-of-flight problems to be solved using Newton's method. The function and its derivatives are more complicated than for orbital position problems, but the method is valid. This method, coupled with a scheme for generating an initial estimate, would be the probable method for implementation in a computer program.

3.5.4 Finding the Orbital Characteristics

In time-of-flight problems, it is usually not just the semimajor axis but all of the characteristics of the orbit that we seek. In Example 3.4 above, for example, knowing the semimajor axis of the cometary orbit is not sufficient to predict, for example, when and where it will reach perihelion. To answer these kinds of questions, we need to find the eccentricity of the trajectory, as well. This knowledge is assumed in the formulation of Kepler's equation, but not so in the derivation of the transit time results. Therefore, we must be able to derive the eccentricity from these results.

Recalling the definitions of α, β, and ξ in section 3.2, we note that

$$\alpha - \beta = 2\kappa = E_2 - E_1 \tag{3.38}$$

and

$$\alpha + \beta = 2\xi \quad \Longrightarrow \quad \cos\frac{\alpha + \beta}{2} = e \cos\frac{E_2 + E_1}{2} \tag{3.39}$$

These two equations are not enough to solve for e. However, recalling that $r = a(1 - e\cos E)$, we can write

$$\frac{\cos E_2}{\cos E_1} = \frac{a - r_2}{a - r_1} \tag{3.40}$$

From (3.38), we can write

$$\cos E_2 = \cos(E_1 + 2\kappa) = \cos E_1 \cos(2\kappa) - \sin E_1 \sin(2\kappa)$$

which we divide through by $\cos E_1$ to get

$$\frac{\cos E_2}{\cos E_1} = \cos(2\kappa) - \tan E_1 \sin(2\kappa) \tag{3.41}$$

We can now include (3.40) in (3.41) to arrive at

$$\tan E_1 = \frac{1}{\sin(2\kappa)} \left[\cos(2\kappa) - \frac{a - r_2}{a - r_1}\right] \tag{3.42}$$

Since the values of α and β are byproducts of the solution of the time-of-flight problem, (3.42) will allow us to find E_1 (and then, should we desire, E_2). We can then use any number of relations to find the eccentricity.

Kaplan [8] attacks the entire problem in a somwhat different matter (though still using the geometry of the transfer). He proceeds to derive a quadratic exression with the semi-latus rectum as the argument. Solving for the case of the ellipse, this provides

$$p = \frac{a(r_1 + d - r_2)(r_2 + d - r_1)}{d^2} \sin^2\left(\frac{\alpha + \beta}{2}\right) \tag{3.43}$$

and, for the hyperbolic problem,

$$p = \frac{-a(r_1 + d - r_2)(r_2 + d - r_1)}{d^2} \sinh^2\left(\frac{\alpha + \beta}{2}\right). \tag{3.44}$$

Taking limiting arguments as in section 3.4, the result for the parabolic case becomes

$$p = \frac{(r_1 + d - r_2)(r_2 + d - r_1)}{d^2} \left(\sqrt{\frac{r_1 + r_2 + d}{4}} + \sqrt{\frac{r_1 + r_2 - d}{4}}\right)^2. \tag{3.45}$$

Between these and (3.42), finding the orbital parameters is relatively straightforward.

Example 3.5. Considering the comet of Example 3.4, how long after the second sighting will the comet reach perihelion, and how far from the Sun will it be?

Solution: By reasoning similar to that used above, it is easy to show that in hyperbolic problems

$$\tanh F_1 = \frac{1}{\sinh(2\kappa)} \left[\cosh(2\kappa) - \frac{a - r_2}{a - r_1}\right] \tag{3.46}$$

From the information given, we compute

$$\alpha = 2.8904 \qquad \beta = 1.8470$$

from which (3.46) gives

$$\tanh F_1 = \frac{1}{1.2434} \left[1.5956 - \frac{-2.686 \times 10^8}{-7.136 \times 10^8}\right] = 0.9806 \quad \Longrightarrow \quad F_1 = 2.3121$$

We can now find the eccentricity from

$$e = \frac{a - r}{a \cosh F} = 1.750$$

From the eccentricity, it is now trivial to compute perihelion.

$$r_p = a * (1 - e) = 60.0 \times 10^6 \text{km}$$

Time to perihelion is found easily from the hyperbolic version of Kepler's equation. Knowing F_1 and e, we solve for time of flight from

$$t - t_p = \sqrt{\bar{a}^3/\mu}\,(e \sinh F - F) = 1.264 \times 10^7 \text{s} = 146.25\,\text{days}$$

Since this is time to perihelion from the first sighting, we subtract the 110 days to the second sighting and find that the comet will reach perihelion in only 36.25 more days.

♠

3.6 Choosing the Right Solution

Although we have derived approaches to the time of flight problem for each of the orbit types, there remains the problem of selecting the appropriate solution. For any problem involving a known focus and two points on the curve, there are two hyperbolic, two parabolic, and four elliptic solutions. We ignored this, tacitly assuming that the solution was straightforward,

It is easily seen that for any two points P_1 and P_2 defined relative to a given focus, there are infinitely many of each conic section that pass through the points. It should not be surprising then that there are multiple possible solutions for a specified travel time. In the hyperbolic and parabolic case, these correspond to solutions that sweep angles of less than or greater than π. For the elliptic case, there are (usually) two possible ellipses that pass through the points, and the two solutions relate again to sweeps of greater than or less than pi.

3.6.1 Parabola and Hyperbola

As mentioned, for parabolic and hyperbolic orbits, the two possible solutions for a given travel time between two points depends upon the sweep angle. If the two points are directly opposed, so that the observations are separated by 180°, the two paths are identical.

For situations in which the observations are separated by an angle less than π, there is no way to know purely from their positions and the travel time which solution is correct. It is necessary to know whether the object has in fact traveled the short path, or taken the longer arc so that the transfer angle is greater than π.

For the parabolic transfer, it has already been mentioned that for transfers of less than π, we use (3.32). For transfers of greater than π, by contrast, we use (3.33). The only difference is changing a subtraction to an addition. For the hyperbolic orbit, the result is effectively the same. Recall that we find the angle β as (equation (3.30))

$$\beta = 2 \sinh^{-1} \sqrt{\frac{r_1 + r_2 - d}{4\bar{a}}}.$$

For transfer angles greater than π, we simply take the negative of this. The subtraction in (3.28) is then a subtraction of a negative, becoming an addition.

3.6.2 Elliptic transfer

In the case of elliptic transfer, the two elliptic solutions each provide two possible transfers, one less than π and one greater. These two will have different travel times, however, so there is still only one solution for each ellipse that satisfies the problem. It remains, however, to find which ellipse to use.

Recall that the choice between elliptic or hyperbolic solution is made by determining the parabolic transfer time. The parabola can be thought of as the *minimum energy hyperbola*, and if the required travel time is greater than the parabolic transfer time, the solution must be elliptic. In the same way, we can determine the *minimum energy ellipse* that satisfies the geometry for a particular problem, neglecting the transfer time. Recall (3.18):

$$\alpha = 2\sin^{-1}\sqrt{\frac{r_1 + r_2 + d}{4a}}.$$

Clearly, there is a minimum value of the semimajor axis, below which the term under radical is greater than unity and the inverse sine has no real solution. This minimum value of a is the semimajor axis of the minimum energy ellipse:

$$a_m = \frac{r_1 + r_2 + d}{4}. \tag{3.47}$$

Using this value of a, $\alpha = \pi$ and $\sin\alpha = 0$, so that

$$t_m = \sqrt{a_m^3/\mu}\left[\pi - (\beta - \sin\beta)\right]. \tag{3.48}$$

If the transfer time is greater than this, it must be true that the transfer is the "long way around" (in time, not necessarily angle swept) on whichever ellipse is the correct solution. Letting α_0 and β_0 be the values of the angles from (3.18) and (3.19), we can specify the possible solutions as

$$\begin{array}{rc} \text{angle swept} < \pi & \beta = \beta_0 \\ \text{angle swept} > \pi & \beta = -\beta_0 \\ \Delta t < t_m & \alpha = \alpha_0 \\ \Delta t > t_m & \alpha = 2\pi - \alpha_0 \end{array}$$

When finding the orbital characteristics using the methods of section 3.5.4, it is necessary to be sure to use the correct values of α and β. Prussing and Conway [12] give a geometric interpretation of these angles, and a more detailed discussion.

3.7 Problems

1. A minor planet has an orbit with eccentricity $e = 0.21654$ and period 4.3856 years. Calculate the eccentric anomaly 1.2841 years after perihelion.

2. Halley's comet has semimajor axis $a = 17.9654$ AU, eccentricity $e = 0.967298$, and periapsis passage was $t_p =$ February 9, 1986. Find the eccentric and true anomalies on October 27, 1960.

3. An object is at perihelion at $r_p = 8.00 \times 10^7$km. When will the object be at $r = 3r_p$ if:

   (a) The orbit is elliptic with $a = 2.5r_p$?

   (b) The orbit is parabolic?

   (c) The orbit is hyperbolic with $e = \sqrt{2}$?

4. Consider the space probe in example 3.3. How far will it be from the Sun when the warranty expires on the first of January, 2003?

5. A satellite in orbit about the Earth has perigee height 622 kilometers and apogee height 2622 kilometers above the surface.

   (a) What is the time of flight from $\nu = 15°$ to $\nu = 120°$?

   (b) What is the time of flight from $\nu = 240°$ to $\nu = 30°$?

6. A comet has an orbit in the plane of the ecliptic. At aphelion the comet's velocity is 9.29 km/s and the distance from the Sun is $r_a = 3.365 \times 10^8$ km. How long after aphelion will the comet cross the orbit of the Earth?

7. An object is observed at an altitude above the Earth of 366 km and 12.5 minutes later at an altitude of 633 km; the angle between the two observations was 45°. Find the semimajor axis and eccentricity of the orbit, and the true anomaly at the time of the first observation.

8. An object is observed at an altitude above the Earth of 7202 km and 14.4 minutes later at an altitude of 11980 km; the angle between the two observations was 80°. Find the semimajor axis and eccentricity of the orbit, and the true anomaly at the time of the first observation.

9. An object is observed at an altitude above the Earth of 92000 km and 2.9 hours later at an altitude of 7062 km; the angle between the two observations was 96°. Find the semimajor axis and eccentricity of the orbit, and the true anomaly at the time of the first observation.

10. A periodic scan of the solar system shows an object at a distance of $r_1 = 1.544 \times 10^8$km from the sun. A scan 38.25 days later, in a direction $50°$ away, shows what appears to be the same object at a distance of $r_2 = 1.602 \times 10^8$km. What is the eccentricity and semi-major axis of the object's orbit, and the true anomalies of the sightings?

11. An object is sighted at a distance of 2.82×10^8 km from the Sun. 90 days later it is sighted at a distance of 3.33×10^8 km, and the angle between the two sightings is $40°$. What are the true anomalies at the two sightings? What is the semimajor axis and the eccentricity of the object's orbit? Is it possible that the object has made a complete orbit between the two sightings?

Chapter 4

Orbital Elements – Putting the Orbit in its Place

In the last chapters, we discussed the orbit of two generic bodies, somewhere in "space". We began with the specification that the bodies existed in some inertial frame, but no mention of any particular frame was later made. We discussed only those characteristics of the orbit that are independent of its inertial characteristics. In this chapter, we consider the orbit in relation to the inertial frame. In a sense, we will be locating the orbit in space.

4.1 Coordinate Frames

The orbital elements that will be introduced later serve, in part, to locate an orbit with respect to an inertial reference frame. For the purposes of definition, we could simply assume the existence of such a frame and go on from there. However, orbital mechanics has been built up from centuries of observation of the motions of the planets their moons, and agreed-upon coordinate frames in which to express these observations have naturally been defined.

There are two frames of particular interest for us. Interplanetary travel will require transfers between planetary orbits, which are sun-centered. Such orbits are called *heliocentric*. The orbits of Earth satellites are defined in the *geocentric*, meaning Earth-centered, frame.

4.1.1 Heliocentric Orbits

The heliocentric coordinate frame has its origin, not surprisingly, at the center of the Sun. The fundamental plane of this frame is the *plane of the ecliptic*, the plane in which the orbit of the Earth lies. The Z-axis is then defined to be normal to this plane, along the angular momentum vector of the Earth.

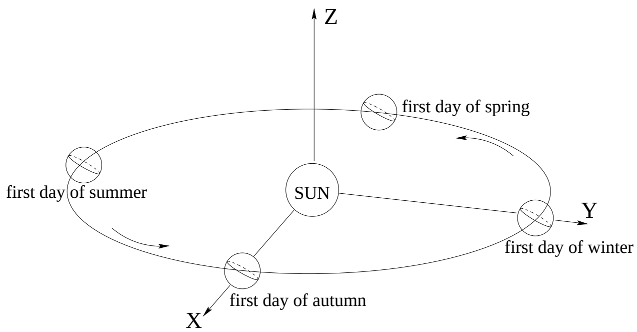


Figure 4.1: Heliocentric coordinate system.

The axes of the heliocentric system are (effectively) fixed with respect to the distant stars. Consider Figure 4.1. The line running from the center of the Earth through the center of the Sun on the first day of spring defines the X-axis. The first day of Spring is the *vernal equinox*, and the X-axis lies in the *vernal equinox direction.* The direction is usually given the symbol ♈ by astronomers, because it used to point to the constellation Aries.[1] The inertial Y-axis is then specified to complete a right-handed orthogonal system.

Many of the terms introduced above come through the ages from very early astronomy, and thus some historical perspective will not be amiss at this point. It will also serve to introduce some further terminology, some of which is still in occasional use. From Berry [3], pp.11-14:

> When the sun crosses the equator the day is equal to the night, and the times when this occurs are consequently known as the **equinoxes**, the **vernal equinox** occurring when the sun crosses the equator from south to north (about March 21st), and the **autumnal equinox** when it crosses back (about September 23rd). The points on the celestial sphere where the sun crosses the equator...are called the **equinoctial points**, occasionally also the equinoxes.
>
> After the vernal equinox the sun in its path along the ecliptic recedes from the equator towards the north.... The time when the sun is at its greatest distance from the equator in the north side is called the **summer solstice**, because then the northward motion of the sun is arrested and it temporarily appears to stand still. Similarly...the **winter solstice**....
>
> Among the constellations which first received names were those through which the sun passes in its annual circuit of the celestial sphere.... This strip of the celestial sphere was called the **zodiac,** because the constellations in it were (with one exception) named after living things (Greek $\zeta\hat{\omega}o\nu$, an animal); it was divided into twelve equal parts, the **signs of the zodiac,** through one of which the sun passed every month.... Owing, however, to an alteration of the position of the equator, the sign Aries, which was defined by Hipparchus in the second century B.C. as beginning at the

[1]Note that the Earth is actually at a negative value of X at this point on its orbit, because the direction dates from when the Earth was considered the center of the celestial sphere. The direction was thus defined as the direction from the Earth to the Sun at the equinox.

> vernal equinoctial point, no longer contains the constellation Aries, but the preceding one, Pisces; and there is a corresponding change throughout the zodiac. The more precise numerical methods of modern astronomy have, however, rendered the signs of the zodiac almost obsolete; but the **first point of Aries** (♈) and the **first point of Libra** (♎), are still the recognized names for the equinoctial points.

(Boldface in the original.) The "alteration of the position of the equator" mentioned by Berry is due to the precession of the axis of rotation (and symmetry) of the Earth. The axis completes one cycle about the celestial Z-axis in about 26 thousand years. For this reason the heliocentric system is not quite fixed, but the motion is sufficiently slow that we may take it as fixed for our purposes. For some long-term or very precise measurements, however, the motion of the equinoxes must be taken into account.[2]

4.1.2 Geocentric Orbits

Consider again Figure 4.1. The Earth's equatorial plane is tilted approximately 23.45° with respect to the plane of the ecliptic. The line of intersection of the two planes will run through the center of both the Earth and Sun on the equinoxes; for the rest of the Earth's orbit it will be parallel to the equinoctial direction. (Another way to define the equinoxes is when the line does pass though the centers of both bodies.) The line of intersection also serves as the X-axis for the geocentric equatorial coordinate system, so that this axis is always parallel to, and in the same direction as, the heliocentric X-axis. The Z-axis is then taken to be the axis of rotation of the Earth, positive northwards in the heliocentric frame, and the Y-axis again is chosen to complete a right-handed system.

4.2 The Classical Orbital Elements

The *elements* of an orbit are a set of parameters that fully describe the orbit, in the sense that given the orbital parameters, it is possible to sketch out the set of points in inertial space that are on the orbit. Another way to think of this is that the elements describe all possible positions of a spacecraft that is somewhere on that orbit. One additional orbital element is added to serve as a reference time, so that the position of that spacecraft can be refined to a specific point as a function of time.

Such a full specification is equivalent to specifying the inertial position and velocity of the spacecraft at any point in time. As it takes three position coordinates and three corresponding velocity components to describe this, it is not surprising that it also requires six elements to fully describe the orbit. These six can be broken down into three subsets: one set (two elements) describes the size and shape of the orbit as

[2]For more on the precession of the equinoxes, see Thomson [15], sec. 5.15, pp. 146 ff. Gurzadyan ([7], chapter 1) has a brief overview of some of the history of Celestial Mechanics.

it lies in the orbital plane; one set (three elements) defines the orientation of the orbit with respect to the inertial reference frame, and the final set (one element) provides the time reference just mentioned.

There are several sets of elements in use in the orbital mechanics community. We will concern ourselves with those that are often referred to as the *classical orbital elements*. These are not in fact the most widely used, due to some mathematical singularities, but they are the most straightforward, and will serve our purposes well. For other sets of elements, the reader is referred to Battin [2], or Danby [6].

4.2.1 Size and Shape of the Orbit

The size and shape of the orbit have been discussed at length in earlier chapters. These are defined by the eccentricity e and the semimajor axis a. Recall that the semi-latus rectum is fully defined by these two, so that any two of the three elements (a, e, p) would give the same information. It is most common to use a and e, though an occasional text will specify orbits through p and e. Because the eccentricity gives the most direct information about the orbit's shape, it is always in the element set.

4.2.2 The Angular Elements

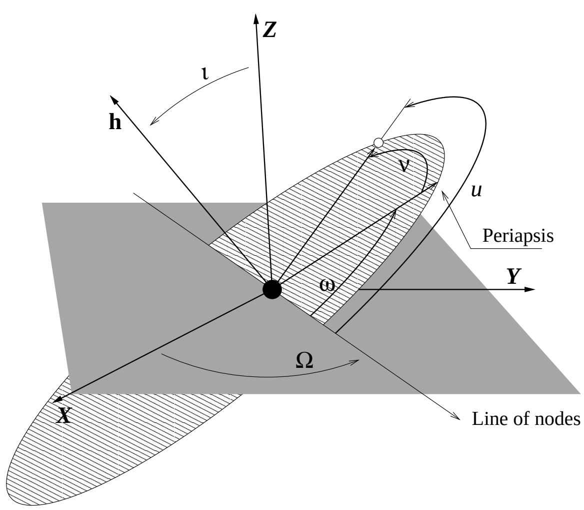


Figure 4.2: Classical orbital elements

Three angular elements are needed to define the orientation of the orbit with respect to the inertial coordinate system. These elements coincide with Euler rotation angles (see Appendix B). They are shown in Figure 4.2, and the names of these elements are:

Ω - The right ascension of the ascending node,

ι - inclination, and

ω - argument of periapsis.

The line of intersection of the plane of the orbit with the fundamental plane of the axis system is known as the *line of nodes*. The nodes themselves are the points at which the orbit itself crosses the fundamental plane. The *ascending node* (sometimes, but not often, the "rising node") is the node at which a spacecraft on the orbit is passing from the southern ($z < 0$) hemisphere to the northern hemisphere; it is "ascending" because $\dot{z}$ is positive. The other node, at which the craft would be going from north to south, is known as the *descending* node.

The term "right ascension" comes from the practice of describing the apparent positions of celestial bodies in terms of the angles describing the line of sight to them in terms of ascension and declination. (These terms are still used in some implementations, but with this one exception will not be needed in this book.) The right ascension of the ascending node is defined as the angle between the inertial X-axis and the line of nodes, measured positive about the positive Z-axis.

The *inclination* of the orbit is the angle between the orbital plane and the fundamental plane. This is also the angle between the inertial Z-axis and the angular momentum vector of the orbit. The angle is measured about the line of nodes, positive by the right-hand rule, with the direction of the line of nodes being taken positive from the descending node towards the ascending node.

The *argument of periapsis* is defined with respect to the Laplace vector, which we recall points from the focus to the periapsis of the orbit. ω is the angle from the line of nodes to the Laplace vector, measured in the direction of travel around the orbit (this is also positive about the angular momentum vector).

4.2.3 Position on the Orbit

There remains the specification of the position of the body on the orbit. This is defined in one of two ways. We can either use the true anomaly at some specified time, or the time of periapsis passage. From either of these bits of information (along with a and e), the position of the body along the orbit at any other known time can be found.

The time of periapsis passage, which we will denote T_p, is simply the time at which the body on the orbit is at periapsis. This is clearly not unique for a closed orbit, since the body will cross periapsis an infinite number of times. For an orbit with non-negative energy, there is only one possible value for T_p.

When choosing to specify the true anomaly at a specified time, the anomaly at *epoch* is usually used. When referring to an observed set of data, the epoch is generally taken to be the time of the observation. For theoretical work, the epoch is an agreed-upon time, set by convention and committee. The current value for solar-system orbits is the precise time of vernal equinox of the year 2000. In 2025, the epoch will be changed to the equinox of 2050. In either case, the symbol we will use for the true anomaly at epoch is ν_0.

4.2.4 Summary

As mentioned, there are many element sets from which to choose. Even within the classical elements, there are many possibilities for inclusion. For our purposes, the set we will normally use will be $(a, e, \iota, \omega, \Omega, \nu_0)$. Note that for the orbit itself, only the first five in the set are needed, so that when it is only the characteristics of the orbit that are of concern the last element might be dropped. Special cases, in which this set may not be valid, are the topic of the next section.

4.3 Special Cases and Singularities

As mentioned, there are many cases in which the classical orbital elements just described are insufficient. This is because there are many common cases in which one or more of them are not defined.

Consider the case of an orbit lying in the fundamental plane (in the geocentric system, this would be called an *equatorial* orbit). In this case, there is no line of nodes, so clearly there can be no right ascension of the ascending node. There can also be no definition of the argument of periapsis, since there is nothing from which to measure it.

There can also be no argument of periapsis if there is no periapsis, which is the case for circular orbits. This is a case of special importance, as many orbits of interest in applications are circular or close to circular.

For these cases, some extensions to the classical element set can be defined. Let u denote the *argument of latitude* of a point along the orbit, defined as the angle between the line of nodes and the ray from the focus to the point. In Figure 4.2, the value of u would be the sum

$$u = \omega + \nu$$

For an inclined circular orbit, this value is still defined.

We will let l denote the *true longitude* along an orbit, which we will define as the sum

$$l = \Omega + \omega + \nu = \Omega + u$$

when these values are defined. For equatorial orbits, l is simply the angle from the X-axis to the line joining the focus to the body on the orbit. Note that this is not true for inclined orbits, because the angle Ω is not in the same plane as the others.

With these definitions, we can define the following additional elements to describe an orbit:

1. Π: The *longitude of periapsis*. This is defined for all non-circular orbits. When the inclination is zero, $\Pi = \Omega$.

2. u_0: The *argument of latitude at epoch*; $u_0 = \omega + \nu_0$. This is defined for any orbit for which the argument of latitude is defined.

3. l_0: True longitude at epoch. $l_0 = \Omega + \omega + \nu_0 = \Pi + \nu_0 = \Omega + u_0$

Although we have been using ν freely, in many cases the position of a body on an orbit is written using l. For general orbits, this is simply ν minus a constant, so that it makes little difference in practice. However, the fact that ν is measured from periapsis can cause troubles in many applications, and so the true longitude is used for safety and full generality.

While the use of any particular element set is a matter of convenience when all elements are constant, this is not the case when the elements can change, as is the case when a spacecraft fires its motors. In the case of a small thrust, we would like the elements of the orbit to change by only a small amount. This is not necessarily the case with the classical elements, in particular when the eccentricity is nearly zero. This case will be explored more fully in the appropriate chapter.

4.4 The Orbital Elements from Position and Velocity

As mentioned earlier, specifying a complete set of orbital elements is the equivalent of specifying the position and velocity of the body on the orbit. It naturally follows that, given the position $\mathbf{r}$ and velocity $\mathbf{v}$ at some specified time, we should be able to compute the elements. In this section, we will show how this is done.

Recall that we assume that the value of the gravitational parameter is known in advance.

The following steps are valid for non-circular orbits. Deriving relationships for circular orbits and other special cases is left as an exercise for the reader. Note that these steps need not be followed in this order, and that this is not the only way to approach the problem.

1. From the vectors $\mathbf{r}$ and $\mathbf{v}$, compute the mechanical properties of the orbit. That is, compute the the specific angular momentum vector, $\mathbf{h} = \mathbf{r} \times \mathbf{v}$, its magnitude h, and the specific energy, $C = \mathbf{v} \cdot \mathbf{v}/2 - \mu/r$.

2. The energy provides the semimajor axis: $a = -\mu/2C$.

3. Compute the Laplace vector $\mathbf{A} = \mathbf{v} \times \mathbf{h} - \mu\mathbf{r}/r$. The eccentricity is then simply the magnitude $\|\mathbf{A}\|$ divided by the gravitational parameter:

   $$e = \|\mathbf{A}\|/\mu.$$

   (Alternatively, $e = \sqrt{1 + 2Ch^2/\mu^2}$, but the Laplace vector will be useful later.)

4. The inclination angle can be immediately computed. Because the inclination is the angle between $\mathbf{h}$ and the inertial Z-axis, we have that

   $$h_3 \triangleq \mathbf{h} \cdot \hat{k} = h \cos \iota \quad \Longrightarrow \quad \iota = \cos^{-1}\left(\frac{h_3}{h}\right)$$

   Recall that the inclination angle is always non-negative and less than or equal to 180°.

5. Find the line of nodes. By definition, the line of nodes lies in the plane of the orbit, and in the fundamental plane. Thus it is normal to both $\mathbf{h}$ and to the Z-axis.

   Now, we know that $\mathbf{a} \times \mathbf{b}$ produces a vector normal to both $\mathbf{a}$ and $\mathbf{b}$. Thus we define the vector $\mathbf{n}$ by
   $$\mathbf{n} = \hat{k} \times \mathbf{h} = -h_2\hat{\imath} + h_1\hat{\jmath}$$
   The order of the cross product is chosen to produce a vector that points towards the ascending node. We need only the direction; we are not concerned with the magnitude. It might however be noted that if $\mathbf{n}$ is normalized to a unit vector at this point, it will be convenient for later computations involving the magnitude n.

6. The right ascension of the ascending node is the angle between X and the line of nodes. Thus
   $$\mathbf{n} \cdot \hat{\imath} = n \cos \Omega \Longrightarrow \Omega = \cos^{-1}\left(\frac{n_1}{n}\right)$$
   If $n_2 > 0$, then Ω is less than π.

7. The argument of periapsis is the angle from $\mathbf{n}$ to $\mathbf{A}$, so we have
   $$\mathbf{A} \cdot \mathbf{n} = An \cos \omega \quad \Longrightarrow \quad \omega = \cos^{-1}\left(\frac{\mathbf{A} \cdot \mathbf{n}}{An}\right).$$
   Recall that, measuring from $\mathbf{n}$, the first half of the orbit is north of the fundamental plane. Thus, if $A_3 > 0$, then $0 < \omega < \pi$.

8. The true anomaly at epoch ν_0 is the angle between $\mathbf{A}$ and $\mathbf{r}$, so
   $$\nu_0 = \cos^{-1}\left(\frac{\mathbf{A} \cdot \mathbf{r}}{Ar}\right)$$
   where we have assumed the reference epoch to be the time of the observation. If $\mathbf{r} \cdot \mathbf{v} > 0$, then ν_0 is less than π.

Naturally, the above recipe fails for circular orbits, and assumes that the element set desired includes ν_0 rather than T_p. However, Kepler's equation can be used to correct for the latter (assuming that the time of the observation is known), while the former requires the use of an appropriate alternate element set.

4.5 From Elements to Position and Velocity

We now know how to find ν, ω, and the other elements, given $\mathbf{r}$ and $\mathbf{v}$ at some time t. The reverse problem is a touch more difficult. The approach requires us to first define one more coordinate frame. It will be simple to express $\mathbf{r}$ and $\mathbf{v}$ as a function of the classical elements in this new frame, and we then use the usual transformations to go from this frame to the inertial.

4.5.1 The Perifocal Coordinate Frame

The *Perifocal* frame is defined by the orbit itself, using the angular momentum and Laplace vectors. The origin is at the focus of the orbit, and the X-axis points to periapsis; from this we get the name.

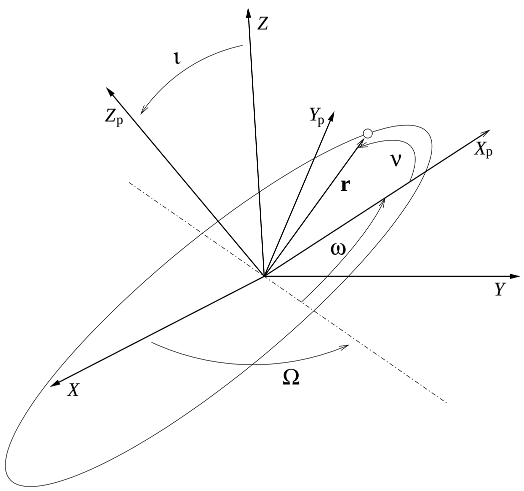


Figure 4.3: The perifocal coordinate frame.

The perifocal system is sketched in Figure 4.3, using the same orbit as in Figure 4.2. The axes of the perifocal system are identified with the subscript p. The easiest mathematical description of the perifocal frame is to say that the X_{p} axis is aligned with the Laplace vector, the Z_{p} axis with the angular momentum vector, and the Y_{p} axis is chosen to complete the right-handed orthogonal system.

Because the Z_{p} axis is aligned with the angular momentum vector of the orbit, it is normal to the orbit plane. Keeping this in mind, and looking at Figure 4.3, it is easy to see that

$$\mathbf{r}_{\mathrm{p}} = r\cos\nu\,\hat{\imath}_{\mathrm{p}} + r\sin\nu\,\hat{\jmath}_{\mathrm{p}}. \tag{4.1}$$

(The notation $\mathbf{r}_{\mathrm{p}}$ here means "the vector $\mathbf{r}$ expressed in the frame subscripted p.")

To find $\mathbf{v}_{\mathrm{p}}$, we differentiate $\mathbf{r}$ to get

$$\mathbf{v}_{\mathrm{p}} = \dot{\mathbf{r}}_{\mathrm{p}} = (\dot{r}\cos\nu - r\dot{\nu}\sin\nu)\,\hat{\imath}_{\mathrm{p}} + (\dot{r}\sin\nu + r\dot{\nu}\cos\nu)\,\hat{\jmath}_{\mathrm{p}} \tag{4.2}$$

Note that the perifocal frame is inertial, so that there is no need to take time derivatives of the unit vectors $\hat{\imath}_{\mathrm{p}}$ and $\hat{\jmath}_{\mathrm{p}}$. To evaluate the derivative, begin with the orbit equation

$$r = \frac{p}{1 + e\cos\nu}.$$

Recalling that $h = r^2\dot{\nu}$ and $p = h^2/\mu$, we differentiate the orbit equation to get (after a fair bit of careful arithmetic):

$$\dot{r} = \sqrt{\frac{\mu}{p}}e\sin\nu \qquad \text{and} \qquad r\dot{\nu} = \sqrt{\frac{\mu}{p}}(1 + e\cos\nu)$$

which we substitute into eqn. (4.2) to arrive at

$$\mathbf{v}_{\mathrm{p}} = \sqrt{\frac{\mu}{p}}\left[-\sin\nu\,\hat{\imath}_{\mathrm{p}} + (e + \cos\nu)\,\hat{\jmath}_{\mathrm{p}}\right] \tag{4.3}$$

Now that we have $\mathbf{r}$ and $\mathbf{v}$ in the perifocal frame, we use our knowledge of coordinate transformations (see B for a review) to put them into the inertial frame.

4.5.2 Perifocal to Inertial Transformation

The angular orbital elements serve the role of Euler rotation angles. To envision this, imagine a rigid wire ellipse in space, of the same size and shape as the orbit, with a wire pointer also running from the focus to the periapsis. Let it begin in the plane of the ecliptic, with the major axis along the inertial X-axis, so that the pointer runs in the positive X-direction. The focus of the ellipse is at the origin of the inertial system.

Now rotate the ellipse about the Z-axis through the angle Ω. This will result in the pointer being along what will be the line of nodes. Next, the ellipse is rotated about its major axis, through the inclination angle i. Finally, rotate the ellipse about its normal (the angular momentum vector of the orbit), through the angle ω. The ellipse is now exactly in the position and orientation of the orbit. This is a standard 3-1-3 Euler rotation.

We let $\mathbf{r}$ be the position vector in the desired inertial frame of reference, and $\mathbf{r}_\mathrm{p}$ the position written in the perifocal frame. The two are related as

$$\mathbf{r}_\mathrm{p} = [\omega \text{ about } Z_\mathrm{p}][\iota \text{ about } X'][\Omega \text{ about } Z]\mathbf{r}$$

which is inverted as

$$\mathbf{r} = [\Omega \text{ about } Z]^T[\iota \text{ about } X']^T[\omega \text{ about } Z_\mathrm{p}]^T\mathbf{r}_\mathrm{p}$$

Filling in the matrices, we have

$$\mathbf{r} = \begin{bmatrix} \cos\Omega & -\sin\Omega & 0 \\ \sin\Omega & \cos\Omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\iota & -\sin\iota \\ 0 & \sin\iota & \cos\iota \end{bmatrix} \begin{bmatrix} \cos\omega & -\sin\omega & 0 \\ \sin\omega & \cos\omega & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{r}_\mathrm{p}$$

which multiplies out to

$$\mathbf{r} = \begin{bmatrix} \cos\omega\cos\Omega - \sin\omega\cos\iota\sin\Omega & -\sin\omega\cos\Omega - \cos\omega\cos\iota\sin\Omega & \sin\iota\sin\Omega \\ \cos\omega\sin\Omega + \sin\omega\cos\iota\cos\Omega & -\sin\omega\sin\Omega + \cos\omega\cos\iota\cos\Omega & -\sin\iota\cos\Omega \\ \sin\omega\sin\iota & \cos\omega\sin\iota & \cos\iota \end{bmatrix} \mathbf{r}_\mathrm{p} \tag{4.4}$$

Being lazy, we write this more compactly as

$$\mathbf{r} = L_{I\mathrm{p}}\mathbf{r}_\mathrm{p} \tag{4.5}$$

Similarly, the velocity vector is transformed as

$$\mathbf{v} = L_{I\mathrm{p}}\mathbf{v}_\mathrm{p} \tag{4.6}$$

and from the last section, we already know $\mathbf{r}_\mathrm{p}$ and $\mathbf{v}_\mathrm{p}$.

4.6 Alternate Element Sets

There is nothing magical about the classical set of orbital elements. They are convenient for many purposes because they have obvious and immediate meanings in terms of the size, shape, and orientation of the orbit. They are not suitable for use in many cases, however.

Any set of six numbers could be used as elements, so long as knowing those numbers would enable us to compute the position and velocity of the body in orbit at any given time. For example, the position and velocity at velocity at some specified time could be used, effectively providing an initial condition for the orbiting body. There are several sets of elements used for varying purposes.

One problem with the classical set is the large number of cases in which some are not defined. Even more troublesome for many uses is the fact that a small change in the velocity of the orbiting body does not necessarily result in a small change in the elements. In particular, consider a satellite in a nearly circular orbit. The argument of periapsis is well-defined. However, a tiny change in velocity may cause the eccentricity to go to zero, and the argument of periapsis is no longer defined. A second small change could make the orbit again elliptical, but the argument of periapsis will have no relation to its earlier value.

While there are a number of alternative element sets, we will here present only two. The first is closely related to the classical set, and has the advantage that it is singular only in very special cases. The second is derived from a different approach to the equations of motion, which we will not explore here.

4.6.1 Equinoctial Elements

The *equinoctial* element set can be defined in terms of the classical set. One of the elements in this set is the semimajor axis a. The other five are defined as

$$L = \varpi + M$$

$$P_1 = e \sin \varpi \qquad P_2 = e \cos \varpi$$

$$Q_1 = \tan\left(\frac{\iota}{2}\right) \sin \Omega \qquad Q_1 = \tan\left(\frac{\iota}{2}\right) \cos \Omega$$

where $\varpi = \Omega + \omega$ is known as the *longitude of periapsis*, and M is the mean motion as defined in section 3.1. This set of elements is singular only for the case of true retrograde orbits, where the inclination $\iota = \pi$.

Note that the use of the mean motion in defining l causes this to no longer be constant. This is somewhat at odds with the description of the elements as a set of six constants. The deviation is due to convenience, as M incorporates both the time since periapsis passage and the mean motion. The element L is usually called the *mean longitude*.

It should be noted that there is no fixed set of symbols for the equinoctial elements.

4.6.2 Delauney Elements

The Delauney elements are derived from the expression of the orbital equations of motion as a Hamiltonian system. This form of mechanics is common in investigations of celestial mechanics, but is not needed for the purposes of this text. The Delauney elements are constants of the motion derived from the Hamiltonian approach, just as we have derived six constants from the Newtonian approach we have taken, again with a modification to include the mean anomoly, bringing time into the element set.

We adopt here Delauney's own notation, which is that most often used for his element set. Defining them in terms of the classical elements, the Delauney set is

$$\begin{aligned} L &= \sqrt{a\mu} & l &= M \\ G &= \sqrt{a\mu(1-e^2)} & g &= \omega \\ H &= \sqrt{a\mu(1-e^2)}\,\cos\iota & h &= \Omega. \end{aligned}$$

Here, M is again the mean anomaly,

$$M = n(t - t_p).$$

The symbols L and h do not have their previous meanings, however. In fact, it is obvious that G is actually the magnitude of the angular momentum, and H is the component along the inertial Z axis.

Note that the Delauney elements are also subject to singularities. The element g is undefined when the eccentricity goes to zero, and L becomes unbounded as the orbit approaches being parabolic. As might be suspected, there is a corresponding set of non-singular elements. As with the equinoctial set, these lack the easily seen physical meaning of most of the Delauney set. They are useful for perturbation analyses in cases when the singularities cannot be confidently avoided, but are well beyond the needs of this text.

We will revisit the Delauney elements briefly in a later chapter, when we discuss the effects of imperfections in the two-body assumption.

4.7 Problems

1. For each of the special cases (a) $e = 0,\ \sin\iota \neq 0$, (b) $e \neq 0,\ \sin\iota = 0$, and (c) $e = 0,\ \sin\iota = 0$, give a complete set of orbital elements, and show how each is computed from $\mathbf{r}$ and $\mathbf{v}$.

2. What are the orbital elements if at some point on an orbit $\mathbf{r} = \hat{\jmath} + 0.2\hat{k},\quad \mathbf{v} = 0.9\hat{\imath} + 0.123\hat{k}$.

3. What are the orbital elements if $\mathbf{r} = (\sqrt{2}/2)\hat{\imath} + (\sqrt{2}/2)\hat{\jmath},\quad \mathbf{v} = (1/2)\hat{\jmath}$.

4. What are the orbital elements if $\mathbf{r} = \hat{k},\quad \mathbf{v} = \hat{\imath}$.

5. Find (by logic and inspection) vectors **r** and **v** such that a circular geocentric orbit has $\Omega = 90°$, $\iota = 45°$, and a period of 2 hours. Explain your selections.

6. In canonical units ($\mu = 1$), the following data is given. Determine a set of independent orbital elements for each case.

   (a)
$$\mathbf{r} = \frac{-3\sqrt{2}}{2}\hat{\imath} - \frac{3\sqrt{2}}{2}\hat{k}$$
$$\mathbf{v} = \frac{\sqrt{6}}{6}\hat{\jmath}$$

   (b)
$$\mathbf{r} = \frac{\sqrt{2}}{2}\hat{\imath} + \frac{\sqrt{2}}{2}\hat{\jmath}$$
$$\mathbf{v} = \frac{-\sqrt{2}}{2}\hat{\imath} + \frac{\sqrt{2}}{2}\hat{\jmath}$$

7. Given the orbital elements: $p = 0.23$, $e = 0.82$, $\iota = \pi/2$, $\Omega = \pi$, $\omega = 260°$, $\nu_0 = 190°$.

   (a) Find **r** and **v** in the perifocal system.
   (b) Convert to the inertial system.

8. Find the inertial position and velocity vectors for the following cases in earth orbit:

   (a)
$$a = 7016\,\text{km}, \quad e = 0.05, \quad i = 45°, \quad \Omega = 0°, \quad \omega = 20°, \quad \nu = 10°$$

   (b)
$$r_p = 6678\,\text{km}, \quad i = 35°, \quad e = 1.5, \quad \Omega = 130°, \quad \omega = 115°, \quad \nu = 0°$$

Chapter 5

Impulsive Orbital Maneuvers

In previous chapters we have examined the two-body problem and discovered that the orbit of a body is constant for all time, at least to the accuracy of the two-body assumptions we used to find the solution. In this chapter, we will consider the problem of changing the orbit by applying thrust to the body in orbit. The change in velocity (in magnitude or direction, or both) may change either or both of the energy and angular momentum, which we have seen control the size and shape or the trajectory.

In this chapter, we will consider only *impulsive maneuvers*. The defining assumption is that the change in velocity occurs over so short a period as the be effectively instantaneous. Because the velocity is assumed to change in zero time, there is no change in position. This is never quite true, but is a close approximation to the case when chemical rockets are used. Rockets generally produce high thrust, and the time spent thrusting is generally so short in comparison with the time of the orbit that we can neglect it for most purposes.

5.1 Initial Considerations

The effect of an impulse is to alter the velocity vector of an orbit by some additive $\Delta\mathbf{v}$. Let the velocity before the impulse be $\mathbf{v}_1$ and the velocity after be

$$\mathbf{v}_2 = \mathbf{v}_1 + \Delta\mathbf{v}$$

Then the angular momentum after the impulse is

$$\mathbf{h}_2 = \mathbf{r} \times (\mathbf{v}_1 + \Delta\mathbf{v}) = \mathbf{r} \times \mathbf{v}_1 + \mathbf{r} \times \Delta\mathbf{v} = \mathbf{h}_1 + \mathbf{r} \times \Delta\mathbf{v} \tag{5.1}$$

so that we see that the change in the angular momentum is zero only if the impulse is zero, or is applied purely in the radial direction. The change in orbital energy also can be zero even for a non-zero impulse, as it is possible to change the direction without changing the magnitude of the velocity.

There are two basic types of problems we will consider in this chapter. One might be thought of as an orbit transition problem, in which we use an impulse to change some aspect of the orbit that a vehicle follows, such as the inclination. This may take only a single impulse, if a point can be found that is on both the initial and the desired orbit. An example of this kind of problem is the inclination change problem, which will be considered later in this chapter. We think of this as a transition problem rather than an orbit change problem because the impulse does not change the initial orbit. Rather, the impulse causes the vehicle to stop following the original orbit and begin following a different orbit, which also passes through the point of application of the impulse.[1]

The other kind of problem we will consider is the *orbit transfer* problem. In this problem, there is no point on the initial orbit that is shared with the desired orbit. An example of this type of problem is raising the orbit of a satellite from low-Earth to geosynchronous orbit. In this case, at least two impulses are required. The first will put the vehicle on a *transfer orbit*, which intersects the initial orbit. The second will be applied at a point on the transfer orbit which is also on the desired orbit, and will serve to transition the vehicle onto that orbit.

It can be shown that any orbit transfer problem can be solved as a two-impulse problem. Such a solution may not be the best solution, of course. It might be better for a number of reasons to have several intermediate orbits in the overall transfer, and to transition from one to another with the needed intermediate impulses.

The usual criterion for "best" solution to an orbit transfer problem is that solution which requires the least total impulse – usually referred to as the least "delta vee". This is because the impulse will generally have to be provided by burning fuel. That fuel first has to be lifted into space, along with tankage for it. We will see in later chapters that each pound lifted into orbit costs many times its own mass in fuel to launch it. For this reason, fuel is a precious resource in space vehicles, and it is not to be expended lightly.

5.2 Single-Impulse Changes

In some cases, the desire is not to transfer from one orbit to a completely different orbit. Instead, many of the parameters of the new orbit may be the same or nearly the same as in the original orbit, with only one or two modified.[2] These maneuvers typically require only a single impulse.

[1] While the difference is purely pedagogical, it serves to make plain the difference between the original orbit (before the impulse is applied) and the new orbit.

[2] It is possible to change all of the elements of an orbit with a single impulse. The only thing that both pre- and post-impulse orbits must have in common is that both pass through the point of application of the impulse.

5.2.1 Change of Inclination

One of the most common of such problems is the need to change the inclination of an orbit. This is a particular problem in the case of satellites intended for equatorial orbits; if they are launched from points away from the equator (such as all launch sites in the United States of America), their initial orbit will of necessity have an inclination at least equal to the latitude of the launch site.[3]

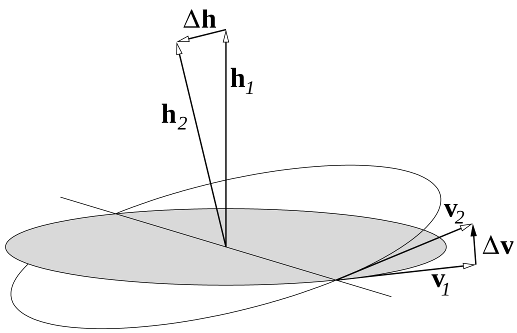


Figure 5.1: Rotating the angular momentum vector.

The inclination of an orbit is defined by the angle between the inertial Z axis and the angular momentum vector $\mathbf{h}$ of the orbit. Thus, changing the inclination is equivalent to rotating $\mathbf{h}$. Recalling eqn. (5.1), we know that

$$\Delta \mathbf{h} = \mathbf{r} \times \Delta \mathbf{v}$$

The discussion here will apply to any rotation of $\mathbf{h}$, which may not result in an equivalent change of the orbital element ι. Thus "plane change", rather than change of inclination, might be a more accurate title.

In general, a plane change might be combined with changes in other orbit parameters, and the necessary impulse will be driven by a combination of factors. However, in the special case of pure plane change of a circular orbit, the needed impulse can be easily computed.

Because the orbit is to have the same energy after the impulse, only the direction of the velocity is to change, not its magnitude. Also, the resulting orbit is to be circular, so that the resulting velocity vector is normal to the radius vector. With these constraints, we can see from geometry that the angle between the initial and final velocity vectors will equal that between the initial and final angular momentum vectors. From the law of cosines, and noting that since $v_1 = v_2$ we can denote both magnitudes simply as v, we have

$$\begin{aligned} \Delta v^2 &= v_1^2 + v_2^2 - 2 v_1 v_2 \cos \Delta\iota \\ &= 2v^2(1 - \cos\Delta\iota) \end{aligned} \tag{5.2}$$

Using the identity

$$2\sin^2\alpha = 1 - \cos(2\alpha)$$

we write this as

$$\Delta v^2 = 4v^2 \sin^2\left(\frac{\Delta\iota}{2}\right) \quad \Longrightarrow \quad \Delta v = 2v\sin\left(\frac{\Delta\iota}{2}\right) \tag{5.3}$$

[3]This is because the launch vehicle does not travel a great distance relative to the size of the Earth during launch, so that it is a good approximation to say that the initial orbit must travel above the launch site. The only way this can happen is for the orbit to have an inclination at least equal to the latitude of the site.

Note that this means that to change the inclination of a circular orbit by 60 degrees would require an impulse equal in magnitude to the velocity of the orbit. Note also that the impulse is not normal to the original orbit plane, because to maintain the same total velocity, the impulse must reduce the tangential velocity as it provides a vertical component. This requires a component directed backwards along the original velocity direction.

This analysis shows that inclination change is a very expensive operation. It explains why it is worthwhile to seek launch points on the equator for a satellite intended for equatorial orbit.

5.2.2 Changes in Orbital Elements

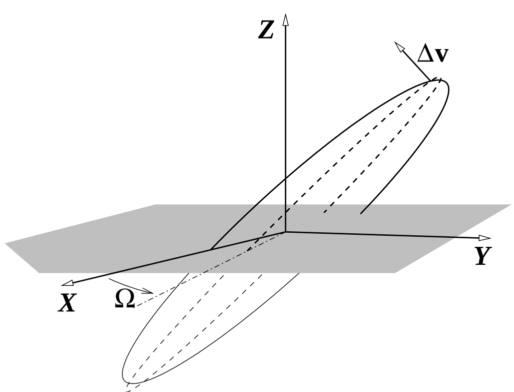


Figure 5.2: Effect of impulse on orbital elements.

Note that the analysis of inclination change above discusses only the rotation of the angular momentum vector about some axis. We did not restrict the point of application of the impulse, which means that the change in inclination achieved is a relative change; that is, the new orbit plane is inclined to the original by $\Delta\iota$. This is not necessarily the change in the *orbital element* ι, as defined in the last section, unless the impulse is applied when the vehicle is crossing the fundamental plane of the inertial axis frame (that is, at the line of nodes).

Consider instead what happens when the impulse is applied ninety degrees from the line of nodes, as illustrated in Figure 5.2. The impulse is applied normal to the plane of the orbit. In this case, there will still be a rotation of the angular momentum, and there is also (unless the impulse is very small) a change in the inclination angle. However, most of the rotation of the angular momentum vector is reflected in a rotation of the orbit plane about the inertial Z-axis, the new orbit will also have different node locations than the original orbit. Some such rotation is likely whenever the impulse is applied at any point other than the line of nodes.

It is worth noting as well that the semimajor axis of the orbit changes, also. This is because the impulse adds energy to the orbit, so that the new orbit has more energy and hence a larger semimajor axis than the original orbit. Thus we see that it is a rare and well-designed impulse that effects only a single element.

5.2.3 A Note on Coordinate Frames

As noted above, changing the plane of the orbit requires rotating the angular momentum vector. This will result in a change in the inclination, but as mentioned in the last section, the change in inclination is not necessarily the same as the rotation of the angular momentum vector. In fact, it would be more accurate to refer to the maneuver as *plane change*, rather than inclination change. The terminology is largely because the usual need for the change is to achieve a desired final orbital inclination, so that the impulse is delivered at the line of nodes and the terms become equivalent.

This points up the need for a coordinate system in which to describe the impulse. As the impulse is defined as the change in a vector quantity, it is itself a vector quantity. In some cases, the coordinate system will be defined by the problem. In others, such as those in the next several sections, it will not. The problem of raising the apoapsis of an orbit, for instance, is independent of the other elements of the orbit, so it can be defined without reference to a global coordinate frame. In general, a global frame will be defined only if the problem requires it. This leaves open the question of defining the impulse vector.

When the coordinate system is not defined, we choose to define the impulse in terms of its radial, tangential, and out-of-plane components. Formally, we define the impulse as

$$\Delta\mathbf{v} = \Delta v_i \hat{\imath} + \Delta v_j \hat{\jmath} + \Delta v_k \hat{k}$$

where $\hat{\imath}$ is in the direction from the orbit focus to the point of application of the impulse, $\hat{k}$ is parallel to the angular momentum vector *prior* to the impulse, and $\hat{\jmath}$ completes the right-handed system.

Example 5.1. A space vehicle is in circular orbit about the earth with inclination $\iota = 40°$, right ascension $\Omega = 45°$, and semimajor axis $a = 7000\,\text{km}$. Compute the impulse needed to increase inclination by 5 degrees, without changing the other orbital elements. What is the point of application of the impulse?

Because the line of nodes is not to change, the impulse must be applied at one of the nodes. Which one does not matter, so we arbitrarily choose the rising node. The point of application is then

$$\mathbf{r} = a\cos\Omega\hat{\imath} + a\sin\Omega\hat{\jmath} + 0\hat{k}$$

There are two approaches to computing the impulse. The first, more complicated, and more generally applicable is to use the techniques of Chapter 4 to compute the velocity at the point of application, for both the pre-impulse and the post-impulse orbit.

The other option is to use aspects unique to this problem to simplify the computations. In this case, the orbit is and remains circular, so that the velocity vector is normal to the position vector before and after the impulse. We also have the magnitude constant across the impulse, with

$$v = \sqrt{\mu_\oplus/a} = 7.546\,\frac{\text{km}}{\text{s}}.$$

Because the point of application is at the rising node, we know that the orbit inclination is the same as the angle between the velocity vectors and the fundamental plane. So the velocity vectors before and after the impulse are

$$\begin{aligned}
\mathbf{v}^- &= v\cos 45^\circ \cos(90^\circ + 45^\circ)\hat{\imath} + v\cos 45^\circ \sin(90^\circ + 45^\circ)\hat{\jmath} + v\sin 45^\circ \hat{k} \\
&= -3.7730\hat{\imath} + 3.7730\hat{\jmath} + 5.3359\hat{k} \\
\mathbf{v}^+ &= v\cos 50^\circ \cos(90^\circ + 45^\circ)\hat{\imath} + v\cos 50^\circ \sin(90^\circ + 45^\circ)\hat{\jmath} + v\sin 50^\circ \hat{k} \\
&= -3.4298\hat{\imath} + 3.4298\hat{\jmath} + 5.7806\hat{k}
\end{aligned}$$

(the superscript "−" and "+" are used to mean "immediately before impulse" and "immediately after impulse", respectively). The impulse is just the first subtracted from the latter:

$$\Delta\mathbf{v} = \mathbf{v}^+ - \mathbf{v}^- = 0.343\hat{\imath} - 0.343\hat{\jmath} + 0.445\hat{k}\,\frac{\text{km}}{\text{s}}.$$

The magnitude of the impulse is $\Delta v = 0.658\,\text{km/s}$. Note that this is the same value we would get using the analysis of section 5.2.1,

$$\Delta v = 2v\sin\left(\Delta\iota/2\right) = 0.658$$

because this is also a pure inclination change problem. The direction of the impulse must be found from this more extensive approach, however. ♠

5.3 Transfer Between Circular Orbits

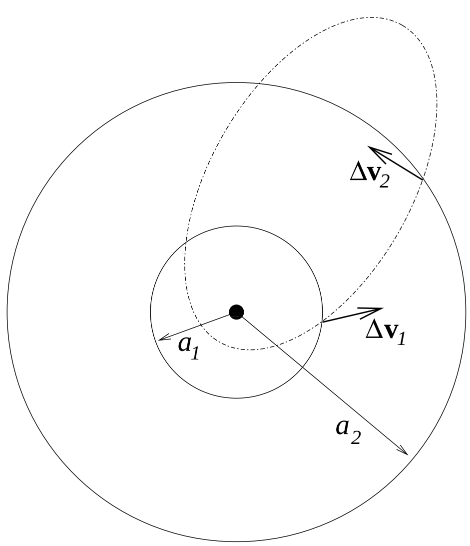


Figure 5.3: Coplanar transfer between circular orbits.

In looking at orbit transfer, we will begin with the most straightforward problem, a two-impulse transfer from one circular orbit to another, with both orbits in the same plane. This transfer will thus require no velocity component normal to the plane of the orbits, and can be considered a two dimensional problem.

Consider the situation as in Figure 5.3. We wish to transfer from the inner orbit with radius a_1, to the outer orbit with radius a_2. We will assume that two impulses are sufficient, one to transfer the craft from the he inner orbit to the transfer orbit, and a second to convert the transfer orbit to the outer circular orbit. Any orbit, such as the one shown in the figure, that intersects both the inner and outer orbits is a candidate for the transfer orbit. Note that in general the transfer orbit need not be elliptic.

As discussed earlier, it is usually desired to find the orbit that requires the least total impulse. In a general coplanar transfer problem, we would have three variables to choose involving the first impulse: the direction of the impulse, its magnitude, and the time of application. Because both orbits in this case are circular, however, the timing has no effect on the geometry of the transfer, and thus no effect on the total impulse required. We therefore ignore it, and have only two variables of concern.

There is no such freedom when it comes to the second impulse, because the point of application is the point at which the transfer orbit crosses the final orbit, and the components of $\Delta\mathbf{v}_2$ are defined by the need to transition to the final orbit. Therefore, a complete mathematical description of the optimization problem would require minimization over two variables. Actually working out the solution requires some effort, but the result is well-known. It was conjectured by Hohmann in 1925, and is referred to as the *Hohmann transfer*. This transfer strategy is known to be the optimal two-impulse transfer between any two circular coplanar orbits.

5.3.1 Hohmann Transfer

The Hohmann transfer is illustrated in Figure 5.4, where the dashed line represents the transfer path between the inner and outer circular orbits. The most notable characteristic of the Hohmann transfer is that the impulse is applied in the same direction as the velocity vector of the circular orbit. The result is a transfer ellipse with periapsis a_1 and apoapsis a_2. A result of knowing the direction of the impulses is that we need only compute their magnitudes, so that the computations become scalar rather than vector equations.

We can compute the total energy required for the Hohmann transfer in a fairly straightforward manner. We know that for any orbit,

$$C = v^2/2 - \mu/r = -\mu/2a$$

Now, circular orbit speed is

$$v_c = \sqrt{\mu/a}$$

and this will give us both the initial and the final velocities. For the transfer orbit, we use the fact that for any ellipse, $2a = (r_p + r_a)$, so that, denoting the semi-major axis of the transfer ellipse as a_t, we have

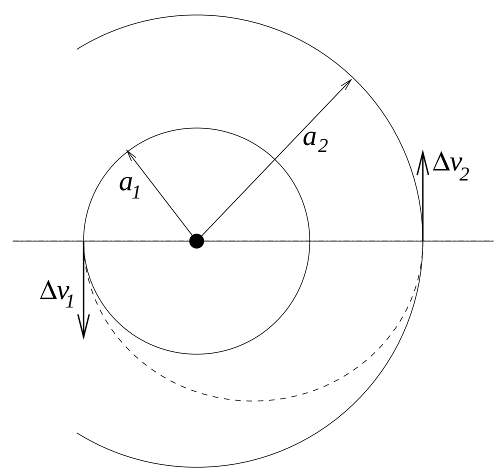


Figure 5.4: Hohmann transfer.

$$a_t = \frac{1}{2}(a_1 + a_2)$$

Then, since the impulses are applied along the direction of the current velocity, the magnitude afterward the first impulse is simply the sum of he circular orbit velocity and the magnitude of the impulse. Likewise, the second impulse is just the amount needed to increase the velocity from that of the apoapsis of the transfer orbit to the desired final circular velocity. We thus have

$$\Delta v_1 = \sqrt{2(C_t + \mu/a_1)} - \sqrt{\mu/a_1} \tag{5.4}$$

$$\Delta v_2 = \sqrt{\mu/a_2} - \sqrt{2(C_t + \mu/a_2)} \tag{5.5}$$

where $C_t = -\mu/2a_t = -\mu/(a_1 + a_2)$. The total impulse required is simply the sum of the two.

As the Hohmann transfer orbit is simply half an ellipse, the transfer time is simply

$$t_f = \pi\sqrt{\frac{(a_1 + a_2)^3}{8\mu}} \tag{5.6}$$

It should be obvious that the transfer from a larger orbit to a smaller follows the same logic. In this case, the only difference is that the impulses are used to slow the vehicle rather than to increase its speed.

Example 5.2. A satellite is carried to low-Earth orbit (altitude 200 km) by the Space Shuttle. Compute the impulses needed to take it into a twelve-hour orbit, using a Hohmann transfer.

Compute the radii of the initial and final orbits. The first is the radius of the Earth plus 200, so $a_1 = 6578.145\,\text{km}$. The second is the radius of a twelve-hour Earth orbit:

$$P = 12 \cdot 60 \cdot 60 = 2\pi\sqrt{a_2^3/\mu_\oplus} \quad\Longrightarrow\quad a_2 = 26610\,\text{km}$$

The energy of the transfer orbit is

$$C_t = \frac{-\mu}{a_1 + a_2} = -12.01\,\text{km}^2/\text{s}^2$$

which allows us to compute its periapsis and apoapsis velocities

$$v_p = \sqrt{2\left(C_t + \frac{\mu_\oplus}{a_1}\right)} = 9.857\,\frac{\text{km}}{\text{s}} \qquad \text{and} \qquad v_a = \sqrt{2\left(C_t + \frac{\mu_\oplus}{a_2}\right)} = 2.437\,\frac{\text{km}}{\text{s}}.$$

The magnitudes of the initial and final impulses are then

$$\begin{aligned}\Delta v_1 &= v_p - v_{c1} = 9.857 - 7.784 = 2.073\,\frac{\text{km}}{\text{s}} \\ \Delta v_2 &= v_{c2} - v_a = 3.870 - 2.437 = 1.434\,\frac{\text{km}}{\text{s}}\end{aligned}$$

(Note: Due to rounding errors, the numbers given in the example may not be completely consistent.) ♠

Note that the Hohmann transfer is the minimum-fuel two-impulse transfer. It is also the two-impulse transfer with the longest time of flight. In many cases, this makes it unsuitable, and we will examine the more general two-impulse coplanar problem in the chapter on interplanetary flight.

5.3.2 Three-Impulse Transfer

While the Hohmann transfer is always the most efficient transfer using only two-impulses, it is sometimes possible to achieve even greater efficiency using three (or more) impulses. The first impulse causes the vehicle to transition from the initial circular orbit to an elliptic orbit; the second impulse then causes transition from this to a second elliptic orbit. This technique is thus often referred to as a *bi-elliptic transfer*.

One particular formulation of this transfer involves using intermediate orbits with very large semi-major axes. This is motivated by the observation that, regardless of the periapsis, the apoapsis velocity (if the eccentricity is sufficiently large), is very small. It is reasonable to conclude that it should therefore require only a very small impulse at apoapsis of a highly eccentric orbit to effect a large change in periapsis.

Consider the limiting case, in which $a_t \to \infty$; that is, the intermediate orbit is parabolic. We will suppose that the impulse is applied in the direction of the velocity, so that we need deal only with the magnitude. The energy of any parabolic orbit is zero, so

$$C_t = \frac{v_1^2}{2} - \frac{\mu}{a_1} = 0 \quad \Longrightarrow \quad v_1 = \sqrt{2\mu/a_1}$$

so that the velocity after the initial impulse is the local escape velocity. The velocity before it is the circular orbit velocity, which we know to be $\sqrt{\mu/a_1}$, so we have that

$$\Delta v_1 = (\sqrt{2} - 1)\sqrt{\mu/a_1}$$

Ignoring the second impulse for now, we note that the third and final impulse will cause the vehicle to go from the return orbit to a circular. Since the return orbit is a very large ellipse, the upper bound on the velocity when the vehicle arrives at the desired orbit is the escape velocity at a_2. By the logic just used, we can then say that the highest possible impulse needed for the transition is

$$\Delta v_3 = (\sqrt{2} - 1)\sqrt{\mu/a_2}$$

The remaining impulse is that required for transition from the first intermediate orbit to the second. Assuming that the impulse is applied at the common apoapsis of the orbits, we have that

$$\Delta v_2 = \sqrt{2\mu\left(\frac{1}{r_a} - \frac{1}{a_2 + r_a}\right)} - \sqrt{2\mu\left(\frac{1}{r_a} - \frac{1}{a_1 + r_a}\right)}$$

where r_a is the apoapsis of the intermediate orbits. As $r_a \longrightarrow \infty$, it is clear that $\Delta v_2 \longrightarrow 0$. In this limiting case, then, the total impulse required for the transfer is

$$\Delta v_{\text{total}} = (\sqrt{2} - 1)\sqrt{\mu/a_1} + (\sqrt{2} - 1)\sqrt{\mu/a_2}$$

If the ratio a_2/a_1 is greater than about 12, this is lower than the impulse required for a Hohmann transfer.

Naturally, it is not possible to make such a transfer, because it would require an infinite amount of time. However, it can be shown that if $a_2/a_1 > 15.582$, a bi-elliptic transfer is always more efficient, so long as the apoapsis of the intermediate orbits is greater than a_2.

5.3.3 Transfers Including Plane Change

It should be no surprise that when plane change is included in the problem, it becomes much more complicated. It is not sufficient to first rotate the orbit plane to the desired inclination, and then perform a Hohmann transfer. Nor is it sufficient to perform the Hohmann transfer first, and then alter the inclination.

Example 5.3. A satellite in a 90-kilometer low earth orbit with inclination 30° is to be sent to an equatorial orbit with period 12 hours. Both initial and final orbits are circular. Compute the necessary impulses if (a) a Hohmann transfer is done first, followed by the plane change, and (b) if the plane change is performed in the same impulse as the terminal thrust of the Hohmann transfer.

Part (a) is simply a matter of summing the Hohmann impulses and the plane-change impulse from section 5.2.1. As the radii of the orbits are the same as in example 5.2, the impulses are the same:

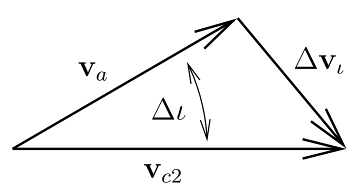
$$\Delta v_1 = 2.073 \quad \text{and} \quad \Delta v_2 = 1.434.$$

The inclination change requires

$$\Delta v_\iota = 2v_{c2}\sin(\Delta\iota/2) = 2 \cdot 3.870\sin(15^\circ) = 2.003$$

so that the total impulse for the transfer plus plane change is $\Delta v_{\text{tot}} = 5.510\,\text{km/s}$.

Combining the last two impulses for (b) requires that the velocity vector on arrival be subtracted from the final vector. The magnitude is



$$\Delta v_\iota^2 = v_a^2 + v_{c2}^2 - 2v_a v_{c2}\cos\Delta\iota \quad\Longrightarrow\quad \Delta v_\iota = 2.141\,\frac{\text{km}}{\text{s}}.$$

This is larger than either the final Hohmann impulse or the plane-change impulse, but is much smaller than the sum of the two. The total for the entire maneuver is $\Delta v_{\text{tot}} = 4.214\,\text{km/s}$, for a savings of $1.296\,\text{km/s}$. ♠

Note that, because the inclination change impulse is proportional to the velocity when it is applied, it would have made no sense to perform the plane change before the transfer. However, while applying it along with the terminal Hohmann impulse is much better, the optimal solution includes dividing the plane change between the two Hohmann impulses, with the majority in the later burn. The actual distribution of plane change depends on the particular problem.

This is also a situation in which the three-impulse maneuver of section 5.3.2 is applicable. Because the plane rotation adds to the total impulse requirement, and is proportional to the velocity, it makes the savings possible even greater. It can be shown that if the required plane change of a circular orbit is greater than about 49°, it is more efficient to use a three-impulse maneuver and make the change along with the second impulse. This is even if there is no increase in the orbit radius.

5.4 More General Transfer Problems

The coplanar circular problem of the last section is one of the few orbit transfer problems that has a simple, well-defined solution. Typically, the solution must be worked out for each individual problem.

Consider a general problem of transfer from an elliptic orbit to a second, inclined elliptic orbit.[4] The variables describing the problem include the sizes and eccentricities of the orbits, their relative inclinations, and the orientations of their semi-major axes relative to their relative line of nodes (that is, the line of intersection of the two orbit planes).

There are three variables in the two-impulse transfer problem in this case. The timing of the initial impulse is arbitrary. The impulse itself has three components, but the impulse must satisfy the constraint that the transfer ellipse intersect the desired final orbit. Because of this, only two of the three components can be freely chosen, giving a total of three free variables. As with the coplanar problem, the particulars of the final impulse will be defined by the velocity vector of the vehicle at the point at which it intercepts the target orbit.

In a three-impulse problem, there are four additional variables, as all of the variables associated with the initial impulse recur for the intermediate impulse. Also, since the intermediate point can be chosen as desired, the constraint on the initial impulse is removed. In such general problems, a three-impulse solution may be desirable even when the outer orbit is not greatly larger than the initial orbit. This is due to the expense of inclination change. Because impulses to change inclination are proportional to the current velocity, it may make little sense to change inclination as part of the initial impulse, as this may be the point of maximum velocity of the transfer orbit.

[4]The orbits need not be elliptic, but the problems are similar for transfer to hyperbolic orbits.

5.4.1 Transfer to Coplanar Elliptic Orbit

Consider what appears to be a very simple problem, as outlined in Figure 5.5. A vehicle is originally on a circular orbit of radius a_0. The desired orbit is coplanar, highly elliptic, and has semi-major axis a. The eccentricity and energy of the target orbit are such that it intersects the original orbit.

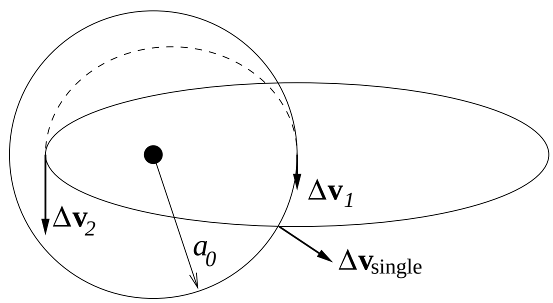


Figure 5.5: Coplanar transfer to elliptic orbit.

In this case, the possible solutions include a single-impulse strategy, as indicated by $\Delta\mathbf{v}_{\text{single}}$ in the sketch. Another possibility is a two-impulse solution. One such is drawn in the sketch, which assumes that the impulses should be planned such that the transfer orbit is an ellipse that is tangent to the target orbit at its periapsis. Note that while there is only one single-impulse solution, there are infinite possible two-impulse solutions.

The single impulse solution is straightforward to compute, given a and the eccentricity e of the target orbit. It is necessary only to compute the velocity vectors of the two orbits at their point of intersection. The velocity before the impulse is

$$\mathbf{v}_0 = \sqrt{\mu/a_0}\,\hat{\jmath}$$

where we define the coordinate system such that $\hat{\imath}$ is radially outward at the point where the two orbits intersect, and $\hat{\jmath}$ is tangent to the circular orbit at that point.

The corresponding velocity on the elliptic orbit is found from the energy and angular momentum of the orbit as

$$v^2 = \frac{2\mu}{a_0} - \frac{\mu}{a}; \qquad h = \sqrt{\mu a(1-e^2)} = a_0 v \sin\phi$$

where ϕ is the included angle between the radius and velocity vectors. From these, we can compute the necessary impulse.

Example 5.4. A spacecraft is on a clockwise circular orbit of radius $a_0 = 2$ (in canonical units). What is the impulse required to transition to a orbit with the same energy, also clockwise, but eccentricity $e = 0.25$?

The velocity on the original orbit is $v_0 = \sqrt{\mu/a_0} = 1/\sqrt{2}$. The magnitude of the velocity after the impulse will be

$$v = \sqrt{\mu\left(\frac{2}{a_0} - \frac{1}{a}\right)} = \frac{1}{\sqrt{2}}$$

(in this case, the velocity does not change, because the impulse does not change the energy). The angular momentum is

$$h = \sqrt{\mu a(1-e^2)} = a_0 v_j \quad \Longrightarrow \quad v_j = 0.6847$$

(note that $v_j = v \sin\phi$). The radial component of the post-impulse velocity is then

$$v_i = \sqrt{v^2 - v_j^2} = 0.1768.$$

The impulse is then

$$\Delta\mathbf{v} = v_i\hat{\imath} + (v_0 - v_j)\,\hat{\jmath} = 0.177\hat{\imath} - 0.023\hat{\jmath}$$

(to three significant digits). ♠

For a two-impulse transfer, it makes sense to consider a Hohmann-like scenario. That is, the first impulse is tangential to the original circular orbit and creates a transfer ellipse that is tangent to the target orbit at either periapsis or apoapsis. This leaves only two cases to analyze.

To periapsis: This is the case illustrated in Figure 5.5. Letting r_p and r_a be the periapsis and apoapsis of the desired orbit, we can compute the characteristics of the transfer orbit as

$$a_t = \frac{a_0 + r_p}{2}; \qquad C_t = \frac{-\mu}{a_0 + r_p}. \tag{5.7}$$

The impulses are then

$$\Delta v_1 = \sqrt{\frac{\mu}{a_0}} - \sqrt{2\left(C_t + \frac{\mu}{a_0}\right)} \tag{5.8}$$

$$\Delta v_2 = \sqrt{2\mu\left(\frac{1}{r_p} - \frac{1}{r_p + r_a}\right)} - \sqrt{2\left(C_t + \frac{\mu}{r_p}\right)} \tag{5.9}$$

Note that because the periapsis of the desired orbit is less than the radius of the initial orbit, the first impulse slows the vehicle. The second impulse will accelerate the vehicle, as the target orbit must have (given the problem statement) more energy than the transfer orbit.

To apoapsis: The analysis for this case is effectively the same. Expressing the energy of the transfer ellipse in terms of a_0 and r_a, the impulses are

$$\Delta v_1 = \sqrt{2\mu\left(\frac{1}{a_0} - \frac{1}{r_a + a_0}\right)} - \sqrt{\frac{\mu}{a_0}} \tag{5.10}$$

$$\Delta v_2 = \sqrt{2\mu\left(\frac{1}{r_a} - \frac{1}{r_a + a_0}\right)} - \sqrt{2\mu\left(\frac{1}{r_a} - \frac{1}{r_a + r_p}\right)} \tag{5.11}$$

In this case, the first burn accelerates, and the second decelerates.

Example 5.5. For the case of example 5.4, compute the total impulse required for transfer to (a) periapis, and (b) apoapsis of the desired orbit.

First, the periapsis and apoapsis of the target orbit are

$$r_p = a(1-e) = 1.5 \quad \text{and} \quad r_a = a(1+e) = 2.5$$

The velocities at periapsis and apoapsis of the transfer ellipse are

$$v_{t,p} = \sqrt{2\left(\frac{-1}{2+1.5} + \frac{1}{1.5}\right)} = 0.8729; \qquad v_{t,a} = \sqrt{2\left(\frac{-1}{2+1.5} + \frac{1}{2}\right)} = 0.6547.$$

The impulses for case (a) are then

$$\Delta v_1 = v_0 - v_{t,a} = 0.0525 \quad \text{and} \quad \Delta v_2 = v_p - v_{t,p} = 0.9129 - 0.8729 = 0.0400$$

for a total impulse of .0925.

For case (b), the velocities at the apses of the transfer ellipse are

$$v_{t,p} = 0.7454 \quad \text{and} \quad v_{t,a} = 0.5963.$$

The impulses are

$$\Delta v_1 = v_{t,p} - v_0 = 0.0382; \qquad \Delta v_2 = v_{t,a} - v_a = .04856$$

for a total of 0.0868.

Note that both of these require much less total impulse than the 0.178 km/s required for the single-impulse approach. ♠

In contrast, consider how much more complicated is the two-impulse approach in full generality.[5] Restricting our discussion to elliptic transfer orbits, we continue the analysis of the last few paragraphs to construct a more general approach. It should be noted that there is no reason to expect that the solution of example 5.5 is not the optimum, and in fact the second is the best two-impulse solution. There are circumstances in which the optimal solution is not feasible, however.

In Figure 5.5, the two-impulse transfer begins with a tangential impulse, and terminates at an apse of the ellipse. In Figure 5.6, the first impulse occurs at point A and puts the craft on a transfer ellipse whose semimajor axis is not aligned with that of the target ellipse, and which crosses the target orbit in two places. Because point A can be chosen anywhere along the initial orbit, and because $\Delta \mathbf{v}_1$ is arbitrary, the characteristics of the transfer ellipse can be chosen to be whatever is desireable. The

[5] To be completely rigorous, we should also include the possibility of out-of-plane motion. This can easily be ruled out in this case, however.

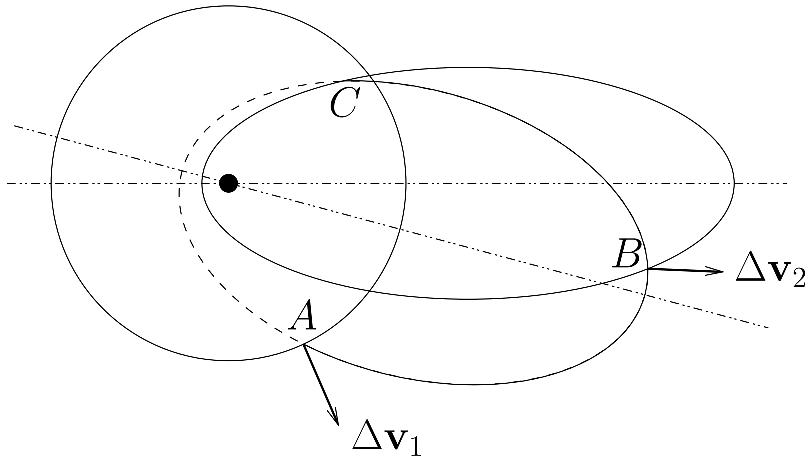


Figure 5.6: General transfer to coplanar elliptic orbit.

mission designer has three variables to choose from, as it is assumed that the impulse out of the plane of the orbits is zero.

The second impulse must be applied when the transfer orbit crosses the target orbit. In Figure 5.6, there are two choices, points B and C. Whichever is chosen, $\Delta\mathbf{v}_2$ will be fully defined by the arrival velocity and the velocity that the craft must have after the impulse to be on the desired final orbit. Computing the point of intersection is not straightforward, and must be done numerically. The time of transfer may then be computed using Lambert's equations.

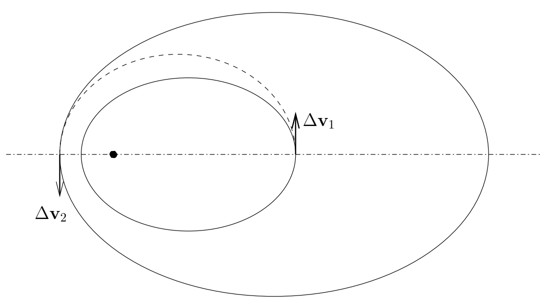


Figure 5.7: Transfer between aligned elliptic orbits.

While the computations are laborious, and the problem has several variables, the problem is not conceptually difficult. The techniques needed to perform the analysis have all been covered in this and previous chapters. However, nothing can be said in general about the solutions to general transfer problems. Each must be solved for the unique conditions of the particular mission requirements. Some relatively simple versions of these problems will be seen later, when interplanetary missions are discussed.

5.4.2 Transfer from Ellipse to Ellipse

The last section discussed a relatively simple problem, because the initial orbit was circular. The general problem is even more difficult when both initial and final orbits are elliptic. We will not explore this problem. However, in the special case of elliptic orbits with colinear major axes, the problem has a fairly straightforward solution.

It can be shown that the minimum two-impulse solution to such problems always involves a transfer from apsis to apsis, as in Figure 5.7. In this case, the transfer shown is from the apoapsis of the inner orbit to the periapsis of the outer orbit. This is not always the case; in fact, is easy to construct situations in which the transfer will be periapsis to periapsis. It depends on the size of the two orbits, and whether the periapses are on the same side of the central body, as in Figure 5.7, or on opposite sides.

5.5 Changes Due to Small Impulse

To this point, we have considered only very large impulses. We will now consider the case of a single small impulse. This is a basic part of *sensitivity analysis*, in which the effect of small changes at one point of a process is considered, in light of the changes they cause later. The discussion begun here can also be considered the simplest investigation of *orbital perturbations*, which is a much broader topic, with a later section of this book devoted to it.

As a simple case, consider the application of some very small impulse to an orbit. To make it even simpler, we will assume that the impulse is applied tangent to the orbit, so that the velocity is changed a small amount in magnitude, and not at all in direction. Beginning with the *vis-viva* equation, we have

$$C = v^2/2 - \mu/r = -\mu/2a \quad \Longrightarrow \quad v^2 = \mu(2/r - 1/a) \tag{5.12}$$

and this is valid for all orbits. We differentiate this expression to get

$$2v dv = \mu \left[\frac{-2dr}{r^2} + \frac{da}{a^2} \right]. \tag{5.13}$$

Since we are assuming a small impulse, we can also assume that the effects will be small.[6] Thus we write δv for dv, et cetera. This leads to

$$2v\delta v = \mu \left[\frac{-2\delta r}{r^2} + \frac{\delta a}{a^2} \right]$$

Here, we use the lower case δ to differentiate this small impulse from the possibly very large Δv that we often compute for maneuvers.

Now, δr is the change in the radius at the time of the impulse and due to the impulse. The impulse is essentially instantaneous, and in any case is an acceleration,

[6] This is the essence of *linearization*, which will be used in detail in later chapters. It is presented more fully in Appendix D.

so that its first-order effects are purely on the velocity. Thus we take δr to be zero, and we have

$$\begin{aligned} 2v\delta v &= \mu\delta a/a^2 \\ &\Longrightarrow \delta a = \frac{2va^2}{\mu}\delta v \end{aligned} \tag{5.14}$$

Consider this a moment. We have said that the radius of the orbit at the point of application of the impulse does not change. The semi-major axis, however, clearly does. This means that if we apply the impulse at perigee, we raise the apogee. Conversely, an impulse applied at apogee raises the perigee.

A more important result of the analysis is that the effect of a small impulse on the semimajor axis is dependent on the velocity at the point of application. Since the velocity is maximized at periapsis, it follows that the best place to apply thrust for the purpose of increasing the semimajor axis is at periapsis. Because a larger value of a is associated as well with greater orbital energy, this is also the optimal point at which to thrust to increast energy. Indeed, beginning with

$$C = v^2/2 - \mu/r,$$

it is easy to show that

$$\delta C = v\delta v. \tag{5.15}$$

Conversely, of course, this also means that any force that causes the spacecraft to decelerate has a greater effect (and drains the orbit of most energy) if applied at periapsis.

5.6 Problems

1. The space shuttle is in a circular orbit of radius R (about the Earth) when it launches a small projectile from its cargo bay. The initial velocity of the projectile *relative to the shuttle* has magnitude u, lies in the plane of the shuttle's orbit, and points outward along the radius vector from the center of the Earth. If $R = 20,000$ km, find:

   (a) The minimum value of u for which the projectile will escape the gravitational field of the Earth.

   (b) The semi-latus rectum of the orbit of the projectile.

   (c) The eccentricity of the orbit in terms of u/v_s, where v_s is the speed of the shuttle.

2. The lunar module (LM) lifts off from the moon and, at burnout, is at periselenium of an elliptic orbit at altitude 30 km above the lunar surface. At apselenium, it will rendezvous with the command module at an altitude of 250 km. The command module is in a circular orbit.

(a) What is the burnout velocity of the LM?

(b) What is the time from burnout to rendezvous?

(c) What impulse is needed to match speeds between the LM and the command module at rendezvous?

3. Compute the impulses needed to go from a low-Earth circular orbit of radius $a = 6578\,\text{km}$ inclined at 28° to a 12-hour circular equatorial orbit for the following cases.

   (a) The orbit is first made equatorial, and then a Hohmann transfer is made to the outer orbit.

   (b) A Hohmann transfer to an inclined 12-hour is performed, and the orbit is then made equatorial.

   (c) The inclination change is combined with the second Hohmann impulse.

4. (a) Compute the impulses needed for a Hohmann transfer from a geocentric equatorial parking orbit at $r_1 = 7000$ km to an orbit at $r_2 = 105000$ km.

   (b) Compute the three impulses needed for a bi-elliptic transfer (with the same initial and final orbits) in which the transfer ellipses have $r_a = 384000$ km

   (c) Compute the time required for each approach.

   Which approach requires the smallest total impulse? What is the time penalty for the bi-elliptic approach?

5. Consider a satellite in an elliptic orbit with eccentricity e and angular momentum h. Calculate the magnitude of the impulse to be applied at point B such that the apse line is rotated 180°, but the eccentricity and angular momentum remain the same. Express your answer in terms of e and h.

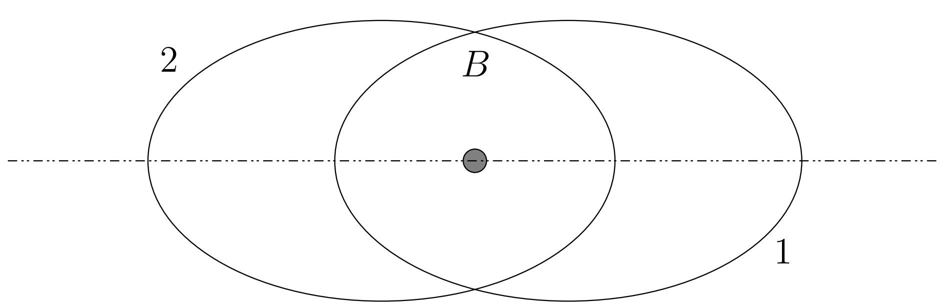


Figure 5.8: Problem 5

6. Compute the minimum Δv required to transfer between two coplanar elliptical orbits which have their major axes aligned. The inner orbit has $r_p = 1.1, e = 0.290$ (canonical units), and the outer has $r_p = 5.0, e = 0.412$. Assume that the periapses of the orbits are on the same side of the focus. You may assume that the transfer will be from apse to apse, and that the impulses are Hohmann-like (that is, they are tangent to the orbit).

7. A spacecraft is in circular equatorial orbit about the earth, with altitude 700 km. Find the impulse required to change the orbit to one with inclination 30°, eccentricity $e = 0.10$, and argument of periapsis $\omega = 270°$. Express the impulse vector in the geocentric coordinate frame. Assume that after the impulse, the rising node will be on the x-axis.

8. Consider a Hohmann transfer from one circular orbit with radius a_1 and inclination ι_1 to a second circular orbit with radius a_2 and inclination ι_2. Part of the required inclination change $\Delta\iota_1$ is done during the first impulse, and the remainder $(\Delta\iota - \Delta\iota_1)$ during the second.

   (a) Find an expression for the total Δv required.

   (b) Find a condition that minimizes the total Δv in terms of $\Delta\iota_1$ (consider the radii of the two orbits, and the initial and final inclinations, as givens in the problem). Show that $\Delta\iota_1 = 0$ is *not* optimal.

9. It is desired to rotate the plane of a circular orbit. Consider the bi-elliptic approach, in which an impulse is first used to transition the craft to a highly elliptic orbit. The plane change is performed at apoapsis of this ellipse, and when the craft gets to periapsis, a transistion is made back to a circular orbit. Find the total Δv needed for this approach, and show that in the limit (as the intermediate orbit becomes nearly parabolic), the bi-elliptic maneuver requires less fuel than a direct rotation for any change greater than approximately 49°.

10. Perform a sensitivity analysis to relate an error in the initial Hohmann impulse to the error in apoapsis radius. If the desired initial impulse is Δv_{nom}, let the actual delivered impulse be

$$\Delta v = \Delta v_{\text{nom}} + \delta v.$$

Define the error in the desired apoapsis radius of the transfer ellipse to be δr_a. Noting that $r_a = 2a - r_p$, find the ratio $\delta r_a/\delta v$ in terms of μ, a_1, and the periapsis velocity v_p of the transfer orbit.

Chapter 6

Interplanetary Trajectories

Having gained an understanding of two-body orbits and basic orbital maneuvers, we are ready to investigate interplanetary trajectories. We will limit ourselves to transfers using high thrust over short duration, such as we expect using chemical rockets. We will also consider only travel within our solar system, though the principles we develop will of course be applicable to any other planetary system we might stumble across in the future.

6.1 Introduction – Patched Conics

The solar system is a very large place, and as we have discussed earlier, it is generally reasonable to consider the planets to be alone within it. When we deal with the orbit of a planet, we concern ourselves only with that planet and with the sun; when dealing with the satellite of a planet, we ignore all bodies other than the satellite and said planet.

In considering interplanetary travel, we will use this idea to divide the trajectory into three phases. In the initial phase, the spacecraft will be very close to the planet of departure, and so we will consider its motion to be defined by the gravitational field of that planet, without concerning ourselves with any other bodies. For this reason, we will consider the motion of the craft relative to the planet, rather than relative to the solar system as a whole. We will place the craft on an escape orbit, thus allowing it to pass from the field of the departure planet into that of the sun.

A primary concept of this approach is that, though the craft has "escaped" the gravitational field of the planet of departure, and thus is at "infinite distance" from that planet, it actually has not moved very far in terms of interplanetary distance. Replace the word "infinite" with "very large". Then consider what "very large" means. We need the distance to be large enough that, when considering the effects of the sun and the planet on the craft's motion, the sun dominates. This leads to the idea of the *Sphere of Influence*, which we will discuss further later. For now, simply accept that the craft can have moved "very far" from the perspective of the planet, while having moved barely at all with respect to the sun.

Once the craft has escaped, we will consider it to be in orbit about the sun. We will note that, though the craft has escape velocity relative to the planet it left, it will generally be in an elliptical heliocentric orbit. We will calculate our escape such that when we find ourselves on this heliocentric orbit, it will be one that intersects the orbit of our target planet, and does so when the target is there.

Finally, upon reaching the target planet, we will either enter parking orbit or execute a planetary fly-by.[1] In either case, we will once again consider ourselves in orbit about the planet, and worry only about the gravitational field of the planet, and motion relative to the planet. We will consider the sun again only if we execute a fly-by and once again enter interplanetary space.

We call this approach to interplanetary trajectories the method of *Patched Conics*, the name deriving from the fact that each section is itself a conic section. The term "patch" comes from the necessity to make the individual sections mate up to form a continuous whole; we do this by "patching" them together.

Before we investigate more fully the patched conic approach, we will discuss the idea of planetary spheres of influence. We will then examine each phase of the trajectory, beginning with the heliocentric, which defines the mission duration and the requirements of the departure phase. A discussion of departure and finally of planetary arrival follows.

6.2 The Sphere of Influence

As mentioned, the solar system is a very large place indeed, and a body can be both "very close" to a planet in one sense, while also being "very far" from the planet from another point of view. In this section, we will get an idea of what "very far" is, at least for the purposes of the motion of the body.

Consider a body in motion near a planet (we will, not surprisingly, consider only the body, the planet, and the sun, ignoring all other planets). We are concerned with the effects on its motion due to the gravitational forces of the sun and of the planet. If the sun dominates the motion, it makes the most sense to write the equations of motion in reference to the sun. If the planet dominates, it makes more sense to write the equations in reference to the planet, and we can ignore the sun. It is this decision that we need to make.

The way we make the decision is to write the equations both ways, and see which one makes the most sense. To see how this is done, we need to go back to the basic mechanics of the system, as we did in investigating the two-body problem, only this

[1]The source of the term "fly-by" should be obvious. We will discuss such trajectories thoroughly in a later section.

time we will have three bodies to consider. To keep our discussion general, we will use masses one and two, rather than planet and sun.

6.2.1 Equations of Motion

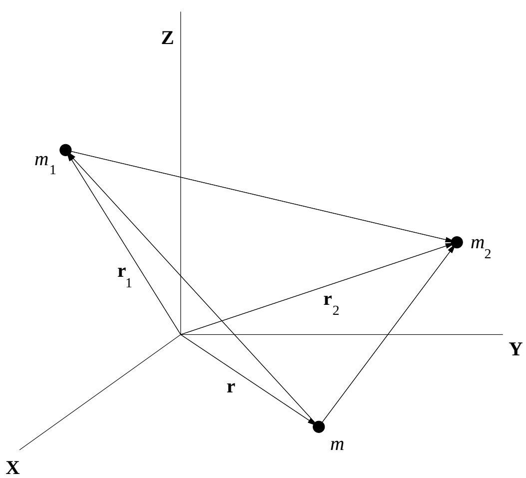


Figure 6.1: Three mass diagram

Consider the situation as in Figure 6.1. We have three masses, m_1, m_2, and m. Each body causes a force on each of the others. The forces are given by

$$\begin{aligned}
\mathbf{F}_1 &= \frac{Gmm_1(\mathbf{r}-\mathbf{r}_1)}{\|\mathbf{r}-\mathbf{r}_1\|^3} + \frac{Gm_1m_2(\mathbf{r}_2-\mathbf{r}_1)}{\|\mathbf{r}_2-\mathbf{r}_1\|^3} \\
\mathbf{F}_2 &= \frac{Gmm_2(\mathbf{r}-\mathbf{r}_2)}{\|\mathbf{r}-\mathbf{r}_2\|^3} + \frac{Gm_1m_2(\mathbf{r}_1-\mathbf{r}_2)}{\|\mathbf{r}_1-\mathbf{r}_2\|^3} \\
\mathbf{F} &= \frac{Gmm_1(\mathbf{r}_1-\mathbf{r})}{\|\mathbf{r}_1-\mathbf{r}\|^3} + \frac{Gmm_2(\mathbf{r}_2-\mathbf{r})}{\|\mathbf{r}_2-\mathbf{r}\|^3}
\end{aligned}$$

Now, we are concerned with the motion of the masses, and thus we use Newton's second law to write, for example,

$$\ddot{\mathbf{r}} = \frac{\mathbf{F}}{m} = \frac{Gm_1(\mathbf{r}_1-\mathbf{r})}{\|\mathbf{r}_1-\mathbf{r}\|^3} + \frac{Gm_2(\mathbf{r}_2-\mathbf{r})}{\|\mathbf{r}_2-\mathbf{r}\|^3}$$

This gives the motion of body m in inertial space. We are interested in the motion with respect to the other two bodies, however. Considering the motion with respect to mass m_1, we are interested in the change in the vector joining them. Thus we write

$$\begin{aligned}
\ddot{\mathbf{r}} - \ddot{\mathbf{r}}_1 &= \frac{Gm_1(\mathbf{r}_1-\mathbf{r})}{\|\mathbf{r}_1-\mathbf{r}\|^3} + \frac{Gm_2(\mathbf{r}_2-\mathbf{r})}{\|\mathbf{r}_2-\mathbf{r}\|^3} - \frac{Gm(\mathbf{r}-\mathbf{r}_1)}{\|\mathbf{r}-\mathbf{r}_1\|^3} - \frac{Gm_2(\mathbf{r}_2-\mathbf{r}_1)}{\|\mathbf{r}_2-\mathbf{r}_1\|^3} \\
&= -\frac{G(m+m_1)(\mathbf{r}-\mathbf{r}_1)}{\|\mathbf{r}-\mathbf{r}_1\|^3} + \frac{Gm_2(\mathbf{r}_2-\mathbf{r})}{\|\mathbf{r}_2-\mathbf{r}\|^3} - \frac{Gm_2(\mathbf{r}_2-\mathbf{r}_1)}{\|\mathbf{r}_2-\mathbf{r}_1\|^3} \qquad (6.1)
\end{aligned}$$

With respect to the second mass, we would have

$$\ddot{\mathbf{r}} - \ddot{\mathbf{r}}_2 = -\frac{G(m+m_2)(\mathbf{r}-\mathbf{r}_2)}{\|\mathbf{r}-\mathbf{r}_2\|^3} + \frac{Gm_1(\mathbf{r}_1-\mathbf{r})}{\|\mathbf{r}_1-\mathbf{r}\|^3} - \frac{Gm_1(\mathbf{r}_1-\mathbf{r}_2)}{\|\mathbf{r}_2-\mathbf{r}_1\|^3} \qquad (6.2)$$

This is all very well, but very theoretical. Let's specialize to the case of a star and a planet, such as in Figure 6.2. We will say that the mass of the star is m_S, that of the planet is m_P, and that of the vehicle m_V. Our question is: Does it make more sense to consider the vehicle to be in motion about the planet, or moving separately from the planet in orbit about the star?

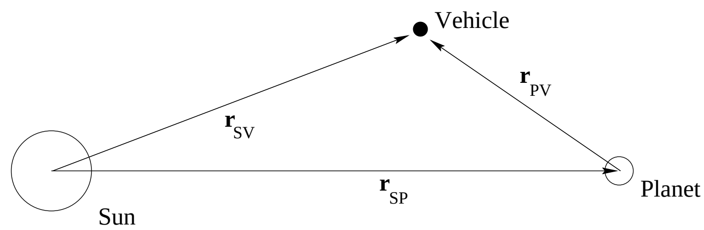


Figure 6.2: Simplified sphere of influence system.

For motion about the star, eqn. (6.2) becomes

$$\ddot{\mathbf{r}}_{SV} = -\frac{G(m_S + m_V)\mathbf{r}_{SV}}{r_{SV}^3} - Gm_P\left(\frac{\mathbf{r}_{PV}}{r_{PV}^3} + \frac{\mathbf{r}_{SP}}{r_{SP}^3}\right) \tag{6.3}$$

and about the planet,

$$\ddot{\mathbf{r}}_{PV} = -\frac{G(m_P + m_V)\mathbf{r}_{PV}}{r_{PV}^3} - Gm_S\left(\frac{\mathbf{r}_{SV}}{r_{SV}^3} - \frac{\mathbf{r}_{SP}}{r_{SP}^3}\right) \tag{6.4}$$

Consider these two equations. The first term on the right hand side of each might be considered the central acceleration term, and were there no further terms, the equations would be identical to the two-body equation, and lead to the standard orbit equation. The second term is the perturbing acceleration. If the central term is much larger than the perturbing term, we can safely ignore the perturbation and treat the system as a two-body system.

6.2.2 The Sphere of Influence

For simplicity, rewrite the equations (6.3) and (6.4) as

$$\ddot{\mathbf{r}}_{SV} = \mathbf{C}_S + \mathbf{P}_S$$

and

$$\ddot{\mathbf{r}}_{PV} = \mathbf{C}_P + \mathbf{P}_P$$

Here, the terms $\mathbf{C}_P$ and $\mathbf{P}_P$ refer to the central and pertubing accelerations relative to the planet[2], and $\mathbf{C}_S$ and $\mathbf{P}_S$ are the same variables referenced to the Star.

We are interested in finding the surface about the planet for which

$$\frac{\mathbf{C}_S}{\mathbf{P}_S} = \frac{\mathbf{C}_P}{\mathbf{P}_P}$$

[2]Note that the perturbing acceleration is *relative* to the planet, and *due to* the presence of the star. The subscript denotes the first point.

Inside this surface, the central term is more dominant in comparison to the perturbing term if the equation is written in terms of motion about the planet. Outside of the surface, the motion is more dominated by the Star, and the planetary attraction serves as the perturbing term.

Tisserand showed in the 1880's that, because the distance r_{SV} will always be much greater than r_{PV} for equality, the surface is very nearly spherical, and its radius is given by

$$r_I = \left(\frac{m_P}{m_S}\right)^{2/5} r_{SP} \tag{6.5}$$

Because of the shape, this surface is known as the *sphere of influence*, also sometimes termed the *sphere of activity*.

As an example, the sphere of influence of the Earth in our solar system is given by

$$r_I = \left(\frac{5.98 \times 10^{24}}{1.99 \times 10^{30}}\right)^{2/5} \cdot 149.5 \times 10^6 = 924,200\text{km}$$

It is important to note that we are not looking for the surface for which the *forces* exerted by the star and the planet are of equal magnitude. For example, at a point directly between the Sun and the Earth on the Earth's sphere of influence, the accelerations due to the Earth and the Sun are

$$\begin{aligned} a_\oplus &= \frac{\mu_\oplus}{r_I^2} = \frac{3.986 \times 10^{11}}{9.242 \cdot 9.242 \times 10^{10}} = 4.67 \times 10^{-4}\,\text{m/s}^2 \\ a_\odot &= \frac{\mu_\odot}{(r_{\odot\oplus} - r_I)^2} = 6.00 \times 10^{-3}\,\text{m/s}^2 \end{aligned}$$

However, these numbers do not include the cetripetal acceleration due to the orbital velocity of the Earth, which almost completely cancels out the attraction of the Sun at this point. This effect must be included when we relate the motion of the vehicle to that of the Earth. When this correction is made, it seems that the Earth dominates the system, rather than the Sun.

It is possible to extend the analysis of this section by defining a small number ϵ and defining two spheres of influence. The inner sphere would be defined as the surface inside of which the effects of the star are much less than those of the planet, satisfying

$$\mathbf{P}_P \leq \epsilon \mathbf{C}_P$$

The outer sphere, similarly, is the surface for which

$$\mathbf{P}_S \leq \epsilon \mathbf{C}_S$$

This analysis is useful in instances in which more accuracy is necessary than is obtained with our more standard definition. This is beyond the scope of our presentation. The interested reader is referred to Roy ([13], pp. 168-171).

6.3 The Heliocentric Phase

An interplanetary mission begins with the departure from the planet of origin. For all missions launched so far, this has of course been the Earth; however, this will possibly not be the case for much longer. Exploratory missions to other planets are being considered that will return samples to the Earth. These return flights will be interplanetary missions starting from other planets, but will be analyzed in the same manner as those from Earth.

At the other end of the mission, the vehicle will enter into the sphere of influence of the arrival planet. Here, either a parking orbit or some sort of fly-by trajectory will be needed. If properly executed, a fly-by trajectory can dramatically increase the heliocentric velocity of the vehicle, sending it onwards to still another planet, or out of the solar system entirely.

Compared to these two phases, where it seems all the action is, the heliocentric trajectory that joins them may seem rather dull. But it is in this trajectory that the vehicle will spend almost all of its time, and it is into another such trajectory that a fly-by would place the vehicle. It is also the requirements of this portion of the mission that define the escape trajectory to be used, and that supply the arrival conditions. We therefore consider this phase first.

6.3.1 Hohmann Transfers

The two main considerations when planning an interplanetary mission are time and fuel. Time is an issue because of the vast distances involved. Fuel is a critical concern because all of the fuel used in the interplanetary phases must first be lifted as payload from the departure planet. This makes each pound of fuel very expensive (both in terms of money and of overall launch mass), and so it is imperative that the mission be planned to require the least fuel possible. Partly this is done by designing the vehicles to be as light as possible, but even more so this is done by choosing trajectories that require the lowest total impulse, in keeping with time constraints.

In considering the heliocentric part of the mission, we will make the simplifying assumption that the orbits of the planets are circular. For most of the planets this a very good approximation. Only Mercury ($e = 0.2056$) and Pluto ($e = 0.2583$) have eccentricities larger than 0.1; the next largest is Mars at $e = 0.0934$. For preliminary examination, assuming circular orbits is good enough, and it greatly simplifies the computations involved. We also assume, for similar reasons, that the orbits lie in the plane of the ecliptic.

Example 6.1. Consider a mission from the Earth to Mars. Assuming circular orbits, we know that the minimum impulse trajectory will be a Hohmann transfer. What is the time of flight on the transfer ellipse?

This is a straightforward application of the Hohmann transfer. The perihelion distance of the transfer ellipse is just the average orbital radius of the Earth, and the aphelion distance is the orbital radius of Mars. Then the period of the ellipse is

$$P = 2\pi\sqrt{\frac{a^3}{\mu}} = 2\pi\sqrt{\frac{(a_\oplus + a_♂)^3}{8\mu_\odot}}$$

The transfer time is half the period, so

$$t_f = P/2 = \pi\sqrt{\frac{(a_\oplus + a_♂)^3}{8\mu_\odot}} = 2.237 \times 10^7 \text{ s}$$

which comes out to 258.9 days. ♠

Note the implicit use of the assumption that the vehicle does not move from a heliocentric standpoint in escaping the Earth. Were that to be taken into account, we would have to allow for variations in both the apses of the transfer orbit, and the fact that the transfer is not precisely half of the Hohmann ellipse. Preliminary time of flight budgeting is also done without considering the escape phase, as the few hours or less in the escape phase is "lost in the noise" compared to the hundreds of days spent in transit.

6.3.2 Non-Hohmann Transfers

For many missions, the Hohmann transfer is sufficient. If we are sending a probe to Venus, the primary consideration will be fuel savings, as there is no harm to the probe in a long transfer time. In other cases, however, there may be restrictions on transfer time, or a particular velocity relative to the arrival planet may be required. Either of these cases would require a non-Hohmann transfer orbit.[3] Non-Hohmann transfers are often referred to simply as "fast" transfers.

Even in cases in which the overall transfer is not Hohmann, we will assume that the spacecraft enters the transfer orbit at perihelion or aphelion, as it would in the Hohmann case. This is to take maximal advantage of the orbital velocity of the departure planet itself.

Example 6.2. It is desired to send a probe from the Earth to Venus in 100 days. What is the eccentricity and aphelion velocity of the transfer orbit?

This is a time-of-flight problem, with known values for radii, but unknown eccentricity and semi-major axis for the transfer orbit. The problem is sketched in Figure 6.3. The object is to find an ellipse such as the dashed one shown such that the aphelion distance of the ellipse is $r_a = a_\oplus$, and the perihelion distance is less than or equal to the orbital radius of Venus. The constraint is that the time of flight from $\nu_1 = \pi$ to

[3]The reasons for a non-Hohmann transfer are not limited to these two examples, of course.

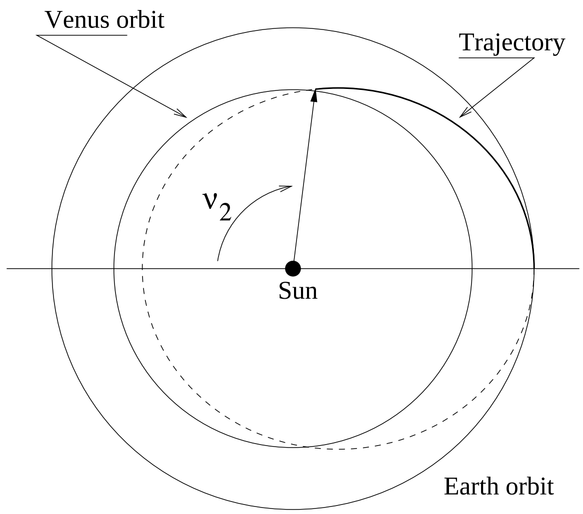


Figure 6.3: Candidate Earth-to-Venus transfer ellipse (Example 6.2).

ν_2, the true anomaly when r of the transfer ellipse equals $a_{♀}$, equals the required t_f of 100 days.

This appears to be a problem such as those solved using Lambert's Theorem in Chapter 3. Those results, however, required that we know not only r_1 and r_2 (the orbital radii of the Earth and Venus in this case), but also the distance between the points on the orbit. Here, we do not have the means to calculate that distance because we do not know the angle between the radius vectors to the two points. Instead, we can solve the problem using iteration on other equations to get the information we need. We will choose a value for a and use that to find the eccentricity. We can then compute the time of flight from ν_2 to perihelion (we will not actually need to compute ν_2 in the process). Subtracting that number from half the period of the transfer ellipse will give the time of flight from aphelion to ν_2, which is the value we are seeking. Depending on the answer, we increase or decrease our guess for a until the time of flight is what we need.

As a starting guess for a, we note that a maximum value is the semi-major axis of the Hohmann transfer ellipse, $a_H = (a_{♀} + a_\oplus)/2$, and a minimum value is $a_\oplus/2$. Since the required time of flight is not very close to the Hohmann transfer time of 146 days, we will try a first guess about halfway between the two values. Letting $a = 1.0 \times 10^8$km, we have

$$e = \frac{r_a}{a} - 1 = \frac{a_\oplus}{a_t} - 1 = 0.49598$$

after which $r = a(1 - e\cos E)$ gives

$$\cos E = \frac{a - a_{♀}}{ae} = -0.2475$$

and finally

$$t_f \quad = \quad P/2 - \sqrt{a^3/\mu}(E - e\sin E)$$

$$= \sqrt{a^3/\mu}\,[\pi - (E - e\sin E)] = 4.507 \times 10^6\text{s} = 52.17\,\text{d}$$

This is far too short a time, so we increase the guess at the semi-major axis to lengthen the flight time. A second guess of $a = 1.15 \times 10^8$ produces the results

$$a = 1.15 \times 10^8 \quad \Longrightarrow \quad e = 0.2313; \quad E = 1.3124; \quad t_f = 6.949 \times 10^6\text{s} = 80.43\,\text{d}$$

A few iterations provides the result

$$a = 1.2155\times10^8\text{km} \quad \Longrightarrow \quad e = 0.1875; \quad t_f = 8.638\times10^6\text{s} = 99.98\,\text{d}$$

♠

Remark 1: As is mentioned in the sections on solving Kepler's equation, there are many well-developed techniques for finding the solutions to iterative systems such as the one in the example. In practice, a computer code implementing such a technique would be used. For illustrative purposes (and for doing homework problems), the method of "informed refinement" (that is to say, intelligent guesswork) is sufficient.

This example is slightly simpler than a corresponding outward transfer, because in this case it was certain that the trajectory would be a piece of an ellipse. This was due to the restriction that the impulse used to begin the transit would be tangential to the orbit of the Earth, so that a parabolic or hyperbolic trajectory would not go inside the orbit of the Earth.

In doing Example 6.2 it was also implicitly assumed that a transfer ellipse existed that satisfied the stated requirements. This is not always the case. It is left as an exercise to find the lower bound on the possible transfer time using this technique. There is no such lower bound if the velocity vector at departure is not restricted in direction.

6.3.3 Phase Angle at Departure

Interplanetary transfers are not just orbit change problems, but *rendezvous* problems. This means that the spacecraft has to arrive at its destination at the same time as some other body, in this case the target planet. This means that it is not sufficient to compute the necessary transfer trajectory. It is also necessary to find the correct time to depart.

Consider the Earth-to-Mars mission of Example 6.1. While the spacecraft is traveling, so is Mars. It is necessary that the spacecraft leave the Earth at a time such that 258.9 days later, when it has reached the orbit of Mars, the planet Mars is at the same place as the spacecraft. This is sketched in Figure 6.4. Measuring angles counter-clockwise, and letting the Earth be at zero at departure, the rendezvous will occur at $\theta_{SC} = \theta_{♂} = \pi$. During the flight, the planet Mars will have moved through an angle equal to 2π times the flight time over the period of the planet's orbit:

$$\Delta\theta_{♂} = \frac{2\pi t_f}{2\pi\sqrt{a_{♂}^3/\mu_{\odot}}} = 2.3677$$

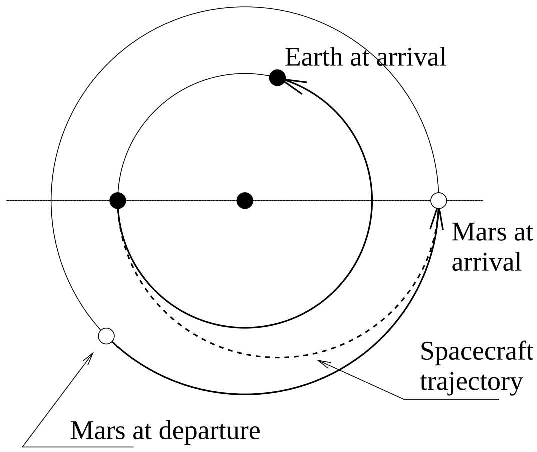


Figure 6.4: Planetary motion during Earth-Mars Hohmann transfer.

From this discussion, it is obvious that the angular motion of any planet during the trajectory is simply

$$\Delta\theta_T = \frac{\sqrt{\mu_\odot}\, t_f}{a_T^{3/2}} \tag{6.6}$$

where the subscript "T" denotes the target planet under consideration.

In the Earth-Mars example, the initial position of Mars is simply $\pi - \Delta\theta_{♂}$, because the angles are being measured in relation to the initial position of the Earth. More generally, the initial position of the target planet in a Hohmann transfer can be written in terms of the initial position of the planet of departure as

$$\theta_{T,i} = \theta_{D,i} + \pi - \Delta\theta_T$$

That this is defined for a Hohmann transfer is evident from the presence of π in the equation as the angle swept by the spacecraft during its trajectory. This is not always the case, as with Example 6.2, where the transfer time is less than the Hohmann transfer. In that case, Eqn. 6.6 can still be used to compute the angular motion of the target planet during the transfer, but the final position must be computed by finding the true anomaly of the spacecraft on its trajectory when it reaches the orbit of the target planet.

The relative angular positions of the planets is often known as the *phase angle*. In the case of the Hohmann transfer, the necessary value at departure is simply given by

$$\theta_{T,D} = \theta_T - \theta_D = \pi - \Delta\theta_T = \pi - \frac{\sqrt{\mu_\odot}\, t_f}{a_T^{3/2}} \tag{6.7}$$

This equation can be written directly in terms of the orbital radii of the departure and target planets by substituting for t_f.

6.3.4 The Synodic Period and Mission Duration

Whatever the angle that is to be swept by the spacecraft during the trajectory, it is clear that the departure and arrival planets must be in the correct relationship at departure for the rendezvous to be successful. Regardless of the phase angle required, the time between occurrences will be the same. This time is known as the *synodic period*.

Consider the phase angle as defined in Eqn. 6.7. The rate of change is computed as

$$\frac{d}{dt}\left(\theta_{T,D}\right) = \frac{d}{dt}\left(\theta_T\right) - \frac{d}{dt}\left(\theta_D\right) = n_T - n_D$$

where n denotes the mean motion of the planet. This is a constant, so the change in the phase angle is

$$\Delta\theta_{T,D} = (n_T - n_D)t$$

The value is positive or negative depending on whether the transfer is inward (to a planet with a smaller orbit and faster mean motion) or outward.

For the planets to regain their initial relative position, the phase angle must go through some multiple of 2π. Since only the absolute value is important, we arrive at the expression for the synodic period

$$T_{\text{syn}} = \frac{2\pi}{\|n_T - n_D\|} \tag{6.8}$$

As an example, the synodic period for Earth-Mars is

$$T_{\text{syn}} = \frac{2\pi}{\sqrt{\mu/a_\oplus^3} - \sqrt{\mu/a_♂^3}} = 6.739 \times 10^7\,\text{s} = 779.9\,\text{d}$$

This period gives rise to the idea of the *launch window*, a period in which the planets are *nearly* in the correct relationship for the transfer. The duration of the window depends on the launch vehicle and spacecraft. All vehicles must have some extra propulsive capability, to overcome errors in calculations, inefficiencies in the engines, and so forth. Some of that extra capability will also allow for a slightly non-optimal transfer. However, should the optimal launch time be missed by too much, the launch window has been missed, and the next opportunity will not come around until the synodic period (minus the length of the launch window) has passed.

The duration of a one-way mission is given by the time of flight. As mentioned earlier, however, some missions currently being planned include a sample return, and of course any eventual manned missions must return the astronauts to Earth. For these missions, a second flight must be planned. Assuming a similar return trajectory, the mission time will be at least doubled. This time is further increased by the need to wait for the planets to come into alignment for the return flight.

The phase angle for the return flight is computed as it was for the initial flight. In the Earth-Mars Hohmann transfer example, the necessary phase angle is given by eqn. (6.7) as

$$\theta_{\oplus ♂} = \pi - \frac{\sqrt{\mu_\odot}\, t_f}{a_\oplus^{3/2}} = -1.311$$

In other words, Mars must be leading the Earth by about 75 degrees for the rendezvous to occur. Since the Earth is in fact leading Mars when the spacecraft first arrives at Mars, there will be a waiting period before the phase angle reaches the necessary value. The methods already outlined can be used to compute the change in phase angle during the first flight, and then the necessary waiting period before the return flight can commence.

The total mission duration for such a mission will include the initial flight to the target planet, the waiting period, and the duration of the return flight. For the Earth-Mars example, the flight time is 258.9 days, the waiting period is 454.3 days, and the return flight will require another 258.9 days, giving a total mission time of about 972 days, or something over two and a half years.

6.4 Departure

Once the characteristics of the heliocentric transfer orbit have been established, it is possible to consider the departure and arrival phases. In departure, the vehicle is given the necessary velocity relative to the departure planet to enter the desired heliocentric orbit.

In the patched conic approach, the departure is considered solely as a trajectory about the departure planet. The restriction is that the velocity vector as $r \to \infty$ *with respect to the departure planet* satisfies the *patch conditions*, which are derived from the characteristics of the heliocentric orbit. This is very similar to the orbit transfer problems considered in an earlier chapter.

6.4.1 Patch Conditions

Once the spacecraft has "escaped" from the departure planet, its velocity is added to that of the planet to provide the velocity vector of the craft with respect to the Sun.[4] In Figure 6.5, the heliocentric velocity vector of a spacecraft ($\mathbf{v}_{SC/\odot}$ – the second subscript will serve to identify the frame of reference for the vector, so that this symbol might be read "the velocity vector of the spacecraft in the heliocentric frame") is seen to be the vector sum of its velocity with respect to the Earth ($\mathbf{v}_{SC/\oplus}$) and the velocity vector of the Earth itself ($\mathbf{v}_{\oplus}$ – it is understood that the velocity vector of a planet will be considered with respect to the heliocentric frame, and so no second subscript is necessary).

Because the sketch is drawn in the heliocentric frame, the spacecraft must be understood to have escaped the Earth, which means it is beyond the sphere of influence. This is an effectively infinite distance when considering the departure trajectory as an Earth-referenced conic section. From a Sun-centered perspective, however, the distance is so small that the craft is considered for all practical purposes to be at the same position as the planet itself.

[4] We will often refer to this as the *heliocentric velocity* or *heliocentric velocity vector*.

Since the craft has escaped, any velocity it has must come from the hyperbolic excess velocity of the departure trajectory. Therefore the magnitude of $\mathbf{v}_{\mathrm{SC}/\oplus}$ is just the hyperbolic excess speed of the Earth-centered escape trajectory. The direction of $\mathbf{v}_{\mathrm{SC}/\oplus}$ is likewise defined by the characteristics of the earth-centered orbit. When added to the velocity of the Earth (or other planet of departure; recall that the Earth is serving here only as an example), the result must be the necessary heliocentric velocity vector to begin the interplanetary transit phase of the mission. It is this necessity to have the terminal conditions of one conic section match the required initial conditions of the next that defines the patched conics approach.

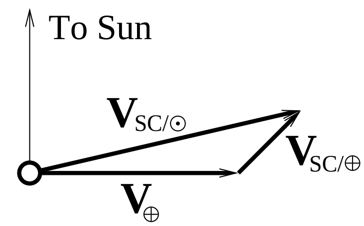


Figure 6.5: Heliocentric velocity of the spacecraft.

Example 6.3. As in Example 6.1, a spacecraft is to go from the Earth to Mars along a Hohmann transfer trajectory. What is the hyperbolic excess speed necessary when the vehicle escapes from the Earth?

As before, we calculate the energy of the Hohmann transfer ellipse as

$$C = \frac{-\mu_\odot}{2a_t} = \frac{-\mu_\odot}{a_\oplus + a_{♂}} = -351.5\,\mathrm{km}^2/\mathrm{s}^2$$

from which we get the perihelion velocity to be

$$v_{SC/\odot} = \sqrt{2(C + \mu_\odot/a_\oplus)} = 32.73\,\mathrm{km/s}$$

Because the transfer ellipse is tangent to the orbit of the Earth, the heliocentric velocity vector of the vehicle as it escapes will be aligned with the orbital velocity of the Earth. Therefore the magnitude of $\mathbf{v}_{SC/\odot}$ is just the magnitude of $\mathbf{v}_{SC/\oplus}$ plus the orbital velocity of the Earth. The latter is found from the usual circular velocity equation:

$$v_\oplus = \sqrt{\mu_\odot/a_\oplus} = 29.78\,\mathrm{km/s}$$

Finally, we have the necessary magnitude of the spacecraft velocity vector to be

$$v_{SC/\oplus} = v_{SC/\odot} - v_\oplus = 2.94\,\mathrm{km/s}$$

(the numbers don't quite add up due to rounding error). ♠

Note that the result of this is precisely the impulse needed for a Hohmann transfer starting from a circular orbit about the Sun with radius $a_\oplus$, ending at an orbit with radius $a_{♂}$. This is effectively what is being done, with the impulses defined in Section 5.3 generalized to the hyperbolic excess speed of the departure trajectory.

As mentioned before, there is no reason to restrict ourselves to Hohmann transfers. Even for fast transfers, however, the departure is generally designed so that

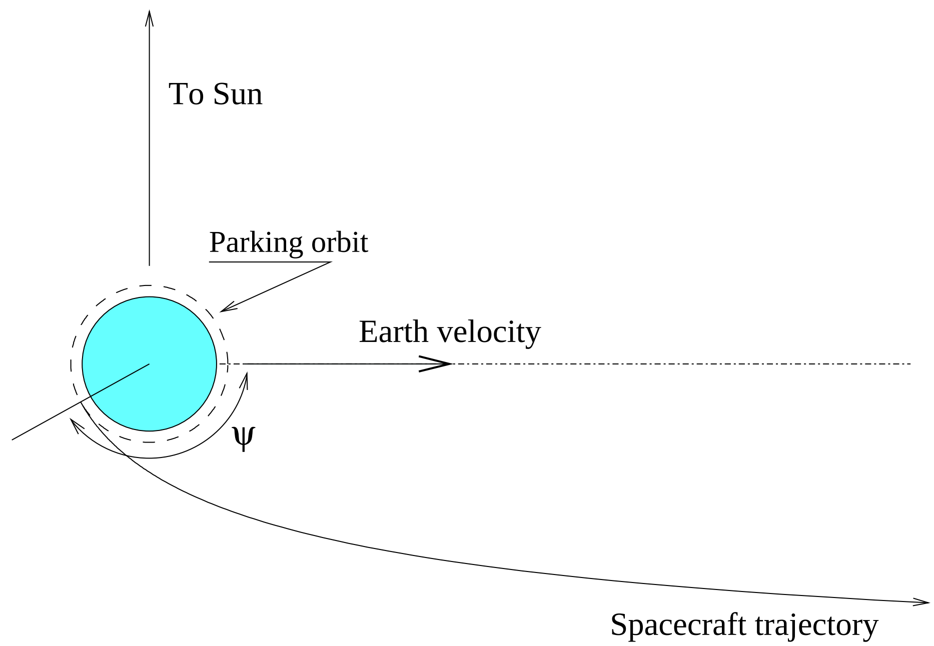


Figure 6.6: Escape trajectory starting from circular parking orbit.

the velocity vector of the spacecraft is aligned with that of the departure planet, in order to take best advantage of the planetary velocity. If this is not the case, the magnitudes of the velocities cannot be added simply, as they are in the example, and vector addition must be employed.

6.4.2 Departure Hyperbola

The departure hyperbola will generally begin from a parking orbit, which we will assume for now to be circular. The impulse needed to accelerate the vehicle and change its trajectory to an escape hyperbola will then be in the same direction as the spacecraft velocity when the impulse is applied. The magnitude and the timing of the impulse must be calculated.

Figure 6.6 shows an escape trajectory from a low-Earth parking orbit.[5] The angle ϕ in the figure measures the point of application of the impulse, in this case referenced to the velocity vector of the Earth. This is a convenient reference, as ψ is then simply the angle of the asymptote for the hyperbolic orbit. We found this earlier as ν_∞, so

$$\psi = \nu_\infty = \cos^{-1}(-1/e) \tag{6.9}$$

[5] The size of the Earth in the figure is reduced; a low parking orbit would not be distinguishable from the circle denoting the Earth were it drawn to scale.

It remains to find the magnitude of the impulse required to convert the circular parking orbit to the required escape hyperbola. This is computed from the hyperbolic excess speed. At the parking orbit radius, the velocity on the escape trajectory is given by

$$C = \frac{v_\infty^2}{2} = \frac{v_1^2}{2} - \frac{\mu_D}{r_1} \quad \Longrightarrow \quad v_1 = \sqrt{v_\infty^2 + \frac{2\mu_D}{r_1}} \tag{6.10}$$

where μ_D is the gravitational parameter of the departure planet and r_1 is the radius of the parking orbit.

At this point, we have made no assumptions about the direction of the impulse. It will be most efficient to apply the impulse tangentially to the parking orbit, to make the best use of the circular orbit velocity, so the velocity after the impulse will then be in the same direction as the velocity before. The magnitude of the impulse is therefore found from simple subtraction as

$$\Delta v = v_1 - v_c = \sqrt{v_\infty^2 + \frac{2\mu_D}{r_1}} - \sqrt{\frac{\mu_D}{r_1}} \tag{6.11}$$

Because the Earth is inclined with respect to the plane of the ecliptic, a parking orbit would have to have an inclination of $-23.45°$ in the usual geocentric frame to have zero inclination in the heliocentric frame. This would result in the departure hyperbola lying in the plane of the Earth's orbit, as sketched in Figure 6.6. However, this is not necessary. Any circular parking orbit that intersects the velocity vector of the Earth can serve as a start for the hyperbola. There is a point on any such orbit at which an impulse can be applied that will result in a hyperbola with the proper energy, and whose asymptotic velocity will be aligned with that of the Earth.

Example 6.4. Continuing the Earth-Mars transfer of previous examples, assume the spacecraft begins in a parking orbit with 200 km altitude. What is the impulse needed for injection onto the escape hyperbola?

The hyperbolic excess speed for injection into the Earth-Mars Hohmann transfer was computed in example 6.3 to be $v_\infty = 2.94\,\text{km/s}$. From eqn. (6.10) we get

$$v_1 = \sqrt{v_\infty^2 + \frac{2\mu_\oplus}{r_1}} = \sqrt{(2.94)^2 + \frac{2\mu_\oplus}{6578}} = 11.396\frac{\text{km}}{\text{s}}$$

The circular orbit velocity of the parking orbit is

$$v_c = \sqrt{\mu_\oplus / r_1} = 7.784\,\text{km/s}$$

and the impulse required for injection is then $\Delta v = v_1 - v_c = 3.611\,\text{km/s}$. ♠

6.5 Planetary Arrival and Capture

The final phase of any interplanetary trajectory is the arrival of the spacecraft at the target planet. Once there, it may be desired to settle into an orbit around the planet, or to use the gravitational field of the planet to modify the heliocentric direction and velocity of the craft. The first of these ideas leads to the problem of *Planetary Capture*.

It is common to misunderstand the term "planetary capture" to mean that the spacecraft is caught in the gravitational field of the planet; in effect "captured". This is not the case. Since the spacecraft is arriving from an effectively infinite distance, it is on a hyperbolic trajectory with respect to the planet of arrival. Such trajectories are not closed orbits, and if nothing is done to change the orbital parameters, the spacecraft will pass by the planet and head back into interplanetary space, its heliocentric trajectory warped by the passage. Such a trajectory is known colloquially as a *Fly-by*. When the fly-by results in some desired change in the heliocentric orbit, it is known more formally as a *Gravity-assist* trajectory.

Thus the problem of planetary capture is that of the spacecraft capturing the planet, rather than vice-versa. And to accomplish this, it is necessary to alter the velocity of the spacecraft with respect to the planet. There are two parts to this process, and we will look at them separately. Whatever else we do, however, we must first translate the heliocentric velocity of the craft to a planet-centered trajectory.

6.5.1 Equating the Heliocentric and Planetary Trajectories

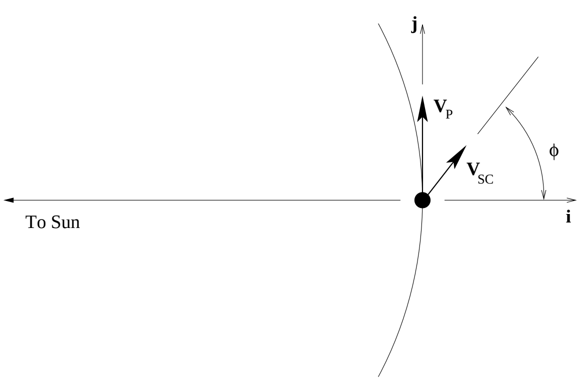


Figure 6.7: Heliocentric view of planetary arrival.

Consider the situation of Figure 6.7. The figure shows the heliocentric velocity vectors of a spacecraft and a planet at the arrival phase of an outward interplanetary transfer. In such a case, the velocity of the craft is typically less than that of the planet. Thus, the spacecraft in fact is ahead of the planet, and the planet must catch up to it as it flies.

The spacecraft is not drawn separately from the planet in Figure 6.7 because from the heliocentric point of view, by the time it is within the planetary sphere of influence it is effectively at the same point as the planet. From the planetary perspective, however, the craft is still at an infinite distance.

To compute the velocity vector with respect to the planet, we note that the heliocentric velocity of the craft can be thought of as the velocity with respect to the planet, plus the velocity of the planet with respect to the Sun. Thus, we have

$$\mathbf{v}_{SC/\odot} = \mathbf{v}_{SC/P} + \mathbf{v}_P \tag{6.12}$$

Since we know the characteristics of the heliocentric orbit, we can find the velocity vector when the spacecraft reaches the planet. We need only define the coordinate frame in which we wish to work.

For the purposes of planetary departure and arrival, a good choice of coordinate frame is that sketched in Figure 6.8. The $\hat{\imath}$ unit vector, which gives the direction of the x-axis, is aligned with the radius vector. The $\hat{k}$ unit vector (the z-axis) is parallel to the angular momentum vector of the orbit of the target planet, and the $\hat{\jmath}$ unit vector is chosen to complete the right-handed orthogonal system. In this system, the velocity vector of the planet (assumed to be on a circular orbit) is purely in the $\hat{\jmath}$ direction, which simplifies computation. The planet itself is at the origin of the frame. Because part of the patched-conic approximation is that the planet-centered trajectories take negligible time, the frame does not move during the approach trajectory. Therefore we consider it inertially fixed for our purposes.

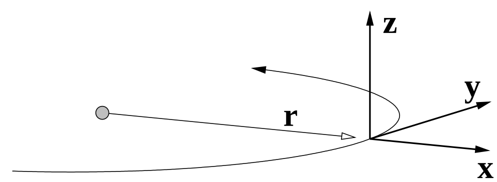


Figure 6.8: Coordinate frame.

Note that the same frame can be used for departure calculations, allowing a more general discussion than was given in Section 6.4.

To compute the spacecraft velocity in this local frame, we need to compute the angle ϕ in the sketch. We can find the value from the specific angular velocity:

$$\mathbf{h}_{SC/\odot} = \mathbf{r}_{SC/\odot} \times \mathbf{v}_{SC/\odot} \Longrightarrow h_{SC/\odot} = r_{SC/\odot} v_{SC/\odot} \sin\phi$$

The value of r is just the radius of the planetary orbit a_P. We know the velocity of the craft from the energy of the transfer orbit. We can thus write the heliocentric spacecraft velocity in terms of its components as

$$\mathbf{v}_{SC/\odot} = v_{SC/\odot} \cos\phi\, \hat{\mathbf{i}} + v_{SC/\odot} \sin\phi\, \hat{\mathbf{j}}$$

The velocity of the craft with respect to the planet is then

$$\mathbf{v}_{SC/P} = v_{SC/\odot} \cos\phi\, \hat{\mathbf{i}} + (v_{SC/\odot} \sin\phi - v_P)\, \hat{\mathbf{j}}$$

Example 6.5. A spacecraft leaves the orbit of the Earth with a heliocentric velocity of 41 km/s, tangent to the Earth's orbit. What is its speed relative to Saturn when it achieves rendezvous?

The energy of the heliocentric orbit is

$$C = \frac{v^2}{2} - \frac{\mu_\odot}{a_\oplus} = \frac{41^2}{2} - \frac{1.32715 \times 10^{11}}{149.6 \times 10^6} = -46.64 \frac{\text{km}^2}{\text{s}^2}$$

The velocity of the craft at the orbital radius of Saturn is then ($a_♄ = 1427 \times 10^6 \frac{\text{km}}{\text{s}}$)

$$v = \sqrt{2\left(C + \frac{\mu_\odot}{a_♄}\right)} = 9.630 \frac{\text{km}}{\text{s}}$$

We compute the specific angular momentum of the orbit as

$$h = r_p v_p = (149.6 \times 10^6) \cdot (41) = 6.134 \times 10^9 \frac{\text{km}^2}{\text{s}}$$

The angle ϕ in 6.7 is then

$$\phi = \sin^{-1}\left(\frac{h}{a_♄ v}\right) = 0.463 \text{ rad} = 26.5^\circ$$

We then have the heliocentric velocity vector of the craft to be

$$\begin{aligned} \mathbf{v} &= v(\cos\phi\,\hat{\mathbf{i}} + \sin\phi\,\hat{\mathbf{j}}) \\ &= 8.617\hat{\mathbf{i}} + 4.300\hat{\mathbf{j}} \frac{\text{km}}{\text{s}} \end{aligned}$$

To convert this to planet-centered coordinates, we need the velocity of Saturn with respect to the sun; this is simply the orbital velocity $v_♄ = 9.654$km/s, and is in the positive $\hat{\mathbf{j}}$ direction. We then re-write eqn. (6.12) to get

$$\begin{aligned} \mathbf{v}_{SC/♄} &= \mathbf{v}_{SC/\odot} - \mathbf{v}_♄ \\ &= 8.617\hat{\mathbf{i}} + (4.300 - 9.654)\hat{\mathbf{j}} = 8.617\hat{\mathbf{i}} - 5.346\hat{\mathbf{j}} \end{aligned}$$

Finally, we have

$$\boxed{v_{SC/♄} = \|\mathbf{v}_{SC/♄}\| = 10.14 \frac{\text{km}}{\text{s}}}$$

♠

6.5.2 Velocity Offset and the Impact Parameter

Now that the velocity vector with respect to the planet has been computed, the characteristics of the planetary orbit can be derived. Whether the spacecraft intends to alter the planet-centered orbit to a parking orbit or continue on, the parameter that defines the problem is the periapsis radius. This is in turn defined by the velocity offset of the of the spacecraft as it enters the sphere of the arrival planet.

Consider Figure 6.9. The spacecraft has arrived at the planetary sphere of influence with some velocity v_∞ with respect to the planet. It is always possible to

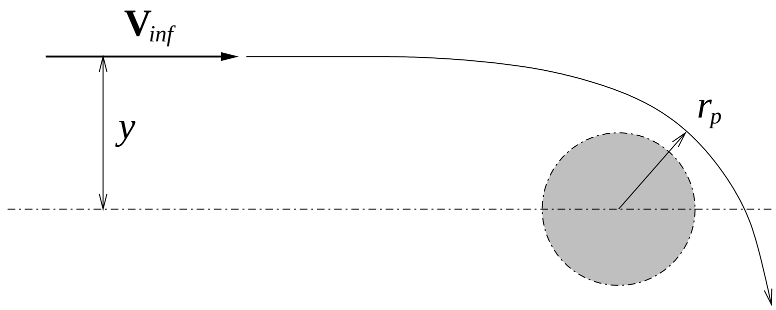


Figure 6.9: Velocity offset at planetary arrival

find a line parallel to the velocity vector that runs through the center of the planet. The distance, denoted y in the figure, between this line and the velocity vector is the offset distance.

From the figure, it is not obvious that this value is very easily changed, and in fact requires a negligible amount of fuel to change. This is because it is not necessary to move the craft to change its offset distance. Instead, it is necessary only to rotate the velocity vector slightly. This will change which line running through the planetary center is parallel to $\mathbf{v}$, and thus change the offset distance.

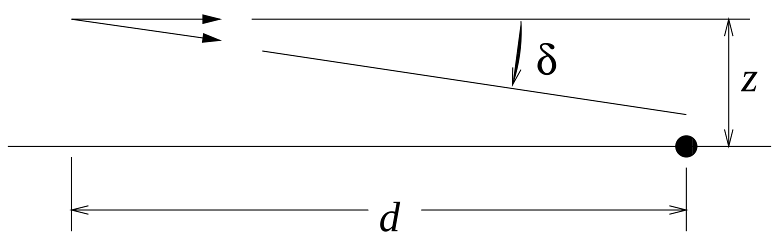


Figure 6.10: Adjusting the miss distance.

To get an idea of the minimal change necessary, consider the geometry of Figure 6.10. Here, the small change δ in the velocity vector has a very large effect on the closest approach to the target point, almost nulling out z. In fact, the angular change to completely null z is given by

$$\delta = \tan^{-1}(z/d)$$

For the situation of planetary arrival, the value of d goes to infinity (in the patched conic approximation) or at the minimum to the radius of the sphere of influence (still extremely large). Thus, we assume that a negligible amount of thrust is needed to give whatever offset we require.

In general, we care about the offset distance only in that it governs the periapsis of the planetocentric trajectory. To compute the periapsis, we will need the specific angular momentum of the trajectory. This cannot be computed using the equations derived earlier directly. Instead, recall that

$$\|\mathbf{h}\| = \|\mathbf{r} \times \mathbf{v}\| = rv \sin\phi = \text{constant}$$

Now, if the distance r is finite, the value of $r\sin\phi$ is just the offset distance y in Figure 6.9. Letting $r \to \infty$ causes $\phi \to 0$, but the product $r \sin\phi$ will remain constant, and the angular momentum becomes[6]

$$h = yv_\infty \tag{6.13}$$

We can relate this to the periapsis radius from

$$h = yv_\infty = r_p v_p \quad \Longrightarrow \quad y = \frac{r_p v_p}{v_\infty}$$

Express the periapsis velocity in terms of r_p and known quantities through

$$\frac{v_\infty^2}{2} = \frac{v_p^2}{2} - \frac{\mu_P}{r_p} \quad \Longrightarrow \quad v_p = \sqrt{v_\infty^2 + 2\mu_P/r_p}$$

Substitution of this result into the preceding expression yields

$$y = \frac{r_p}{v_\infty}\sqrt{v_\infty^2 + \frac{2\mu_P}{r_p}} \tag{6.14}$$

By squaring both sides of this equation, we can derive a quadratic equation for the periapsis radius:

$$r_p^2 + 2\mu_P r_p/v_\infty^2 - y^2 = 0$$

A more immediately useful result comes by setting the periapsis radius r_p to the radius of the planet r_P in eqn. (6.14). The resulting offset distance is known as the *impact parameter*, and is the minimum offset needed to avoid collision with the planet. Denoting this as b, we have

$$b = \frac{r_P}{v_\infty}\sqrt{v_\infty^2 + \frac{2\mu_P}{r_P}} \tag{6.15}$$

6.5.3 Planetary Capture

As mentioned above, planetary capture does not refer to a wandering body being somehow snared by a planet's gravitational field. In the patched-conic approximation, this cannot happen. Instead, the problem we address is that of converting the

[6] Note that this is the standard expression "angular momentum is the velocity times the perpendicular distance" familiar to all engineering and science freshmen. This should not be surprising; although $\mathbf{h}$ was derived originally as a constant of integration, its physical meaning is clear.

hyperbolic approach trajectory into a closed orbit about the planet. This is very much like the problem of departure in reverse.

Consider the problem of injecting into a circular parking orbit at the periapsis of the approach hyperbola. The equations used in this case are in fact the same as for the departure.

Example 6.6. Consider the approach trajectory defined in example 6.5. What impulse is required to put the vehicle into a circular parking orbit of 5,000 km altitude?

Recall that we have assumed that we can choose the periapsis of the approach hyperbola for negligible cost. We therefore assume that these adjustments have already been made, and we may assume that the periapsis will occur at the required altitude. The distance from the center of Saturn at periapsis will then be

$$r_p = 5000 + r_{♄} = 65000\,\text{km}$$

The velocity of the spacecraft relative to Saturn was computed in example 6.5 to be $v_{sc/♄} = 10.14\,\text{km}$. Because the spacecraft is still effectively an infinite distance from the planet when this calculation is made (at least from the planet-centered viewpoint), this is the velocity at infinity for the hyperbola about Saturn. The energy is then

$$C = v_\infty^2/2 = 51.42\,\text{km/s}$$

The velocity at periapsis is found by equating the energy at $r \to \infty$ and r_p:

$$\frac{v_\infty^2}{2} = \frac{v_p^2}{2} - \frac{\mu_{♄}}{r_p} \quad\Longrightarrow\quad v_p = \sqrt{v_\infty^2 + \frac{2\mu_{♄}}{r_p}} = 35.64\,\frac{\text{km}}{\text{s}}$$

Note that this is precisely eqn. (6.10). The necesary impulse to slow the vehicle and inject it into the parking orbit is just the difference in the periapsis and circular orbit velocities.

$$\Delta v = v_p - \sqrt{\mu_{♄}/r_p} = 35.64 - 24.16 = 11.48\,\text{km/s}$$

Note that this could be done directly with eqn. (6.11). ♠

The impulse is found from scalar arithmetic in the example because we assumed that it would be applied at periapsis. This would be done so that the impulse is used most efficiently, and is guaranteed to be possible due to the ease with which the periapsis radius can be adjusted while the spacecraft is still far from the planet. Again, there is no reason why an impulse could not be applied at some other point on the approach hyperbola, but this would waste fuel (unless it were done to correct an emergency, of course).

We also assume that the inclination of the orbit about the target planet is chosen through the orientation of the approach. Because the vehicle can be aimed to pass over the equator, the poles, or anywhere in between, the choice of parking orbit is wide. The only restriction is that the orbit will intersect a line parallel to the direction of approach and passing though the center of the planet.

6.6 Gravity Assist Trajectories

The colloquial term for a spacecraft flight that skirts a planet is for obvious reasons a *fly-by*. A main reason for such trajectories is to obtain information about the planet, and such opportunities are rarely wasted. From the orbital mechanics standpoint, however, the utility of these orbits is the ability to use the planets gravity to alter the heliocentric orbit. In so doing, it is possible to alter not just the direction but also the energy of the trajectory, gaining several kilometers per second of velocity at little or no cost to the vehicle.[7] The use of the gravitational field of the planet to accomplish this leads to the term *Gravity Assist*.

6.6.1 Planar Fly-bys

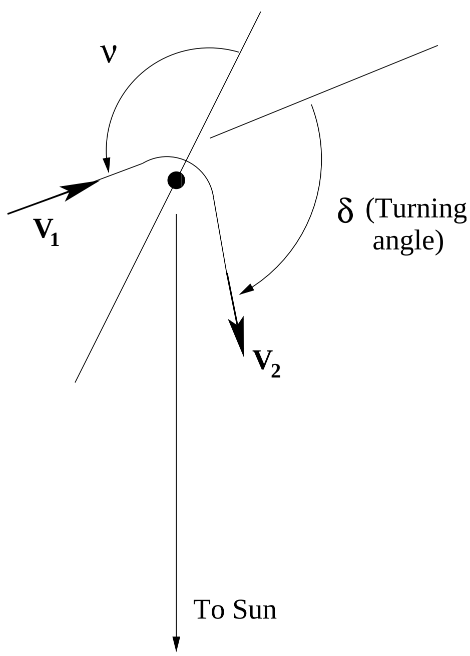


Figure 6.11: Geometry of a fly-by trajectory.

The situation of Figure 6.11 shows a spacecraft on a hyperbolic trajectory past a planet. The figure is drawn from the planetary reference frame, so that the planet itself does not appear to move during the fly-by. The vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in the figure are the velocities of the spacecraft before and after the fly-by, respectively. The angle δ is the *turning angle*; the velocity vector rotates through this angle during the fly-by. Recall that because energy is conserved throughout the fly-by, the magnitude of the velocity (with respect to the planet) does not change.

To compute δ, we note that

$$\delta + 2(\pi - \nu_\infty) = \pi \quad \Longrightarrow \quad \frac{\delta}{2} = \nu_\infty - \frac{\pi}{2}$$

Since we already know that for a hyperbolic trajectory $\cos\nu_\infty = -1/e$, and for any angle α, $\cos(\alpha - \pi/2) = -\sin\alpha$, this leads us to

$$\sin\left(\frac{\delta}{2}\right) = \frac{1}{e} \tag{6.16}$$

At this point, it is necessary to assume some additional information. The energy of the orbit is defined by the magnitude of the velocity, but the eccentricity is defined

[7]There is a cost, of course. The momentum gained by the spacecraft is balanced by an equal loss in the momentum of the planet. We ignore this as negligible.

by the periapsis radius. Since we saw in the last section that this is easily controlled by the spacecraft, we will assume for now that r_p is either given, or is to be computed from other given information (for example, the desired turning angle may be specified, and from this r_p can be computed).

To relate r_p to the eccentricity, recall

$$r_p = \frac{p}{1+e} = \frac{h^2/\mu}{1+e} = \frac{r_p^2 v_p^2}{\mu(1+e)} \quad \Longrightarrow \quad e = \frac{r_p v_p^2}{\mu} - 1$$

Noting that $v_p^2 = 2(C + \mu/r_p)$, this gives

$$\begin{aligned} e &= \frac{2r_p(v_\infty^2/2 + \mu/r_p)}{\mu} \\ &= \frac{r_p v_\infty^2 + 2\mu}{\mu} - 1 \\ &= \frac{r_p v_\infty^2}{\mu} + 1 \end{aligned} \tag{6.17}$$

Thus, given δ, the necessary value of r_p can be computed.

The value of the gravity-assist trajectory lies in the heliocentric orbit that results from the fly-by. The energy of the planetocentric orbit is constant with respect to the planet during the fly-by, but this does not imply that the energy of the resulting heliocentric orbit is the same before and after. In Figure 6.12, the rotation of the spacecraft heliocentric velocity vector is shown. The dashed vectors represent the values before the fly-by, and the solid vectors after (the velocity of the planet is of course not changed).

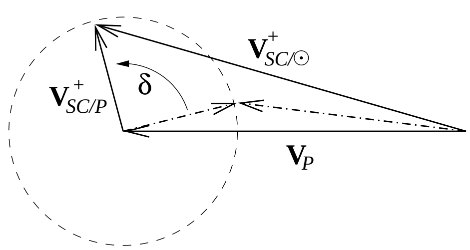


Figure 6.12: Rotation of spacecraft velocity vector during fly-by.

Since the heliocentric velocity of the craft after the fly-by is simply the sum of the velocity with respect to the planet and the velocity of the planet itself, the result will be a vector whose tip lies on the circle in Figure 6.12. If $\mathbf{v}_{SC/P}$ before the fly-by is nearly in the opposite direction of $\mathbf{v}_P$, the resulting rotation will likely result in a heliocentric vector with greater length, since the origin of the vector does not change. For a Hohmann transfer, the spacecraft velocity vector at arrival is aligned with the velocity of the planet. Thus for an outbound Hohmann transfer, for which $\mathbf{v}_{SC/P}$ is in the opposite direction of $\mathbf{v}_P$ in the sketch, *any* fly-by will result in a greater heliocentric velocity for the spacecraft. The situation for an inbound Hohmann transfer is precisely the opposite; since the spacecraft velocity before the transfer is in the same direction, any fly-by will result in a heliocentric orbit with less energy. This may still be desirable as a way to produce large rotations in the vector.

In general, the spacecraft can be expected to be going slower than the planet. As can be seen in the sketch, a fly-by that passes on the sunward side of the planet and then behind it will produce a counter-clockwise rotation of the spacecraft velocity, and will thus increase the orbital energy. A fly-by in front of the planet produces a clockwise rotation and a reduction in the energy. This is not always the case, of course; a large clockwise rotation of the velocity may still increase the energy, if the angle between $\mathbf{v}_{SC/P}$ and $\mathbf{v}_P$ was originally close to π.

Example 6.7. Consider the situation of Example 6.5. The view from Saturn might be as seen in Figure 6.13. Here, $\mathbf{v}_i$ is the initial velocity vector, and $\mathbf{v}_f$ is the final vector, as the spacecraft escapes after the heliocentric fly-by. The value of the angle θ in the figure is easily computed as

$$\theta = \tan^{-1}\left[\frac{\|\mathbf{v}\cdot\hat{\mathbf{i}}\|}{\|\mathbf{v}\cdot\hat{\mathbf{j}}\|}\right]$$

Suppose that we desire a fly-by trajectory such that the out-bound heliocentric velocity is unchanged, but the sign of the $\hat{\mathbf{j}}$ component is reversed (note: this is purely for illustrative purposes; there is no general reason to perform such a maneuver). Find the final heliocentric velocity vector and the energy of the new heliocentric orbit.

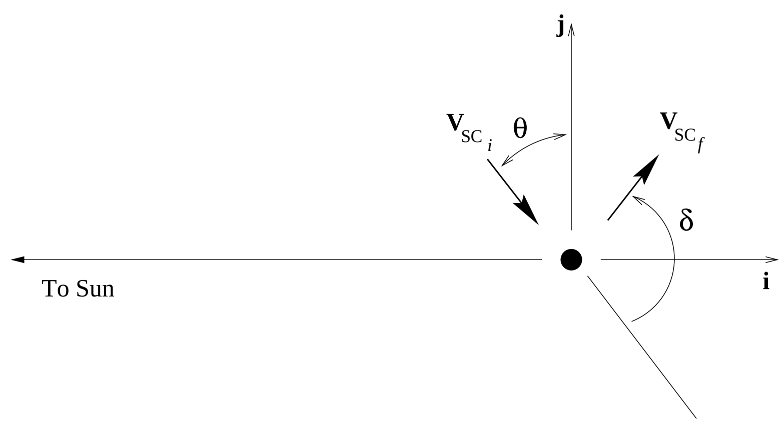


Figure 6.13: Fly-by from the planetary point of view

First, it is important to note that the magnitude of the heliocentric velocity vector will change; the problem does not require that the magnitude of the $\hat{\mathbf{j}}$ component of the heliocentric velocity vector be the same, only that it has a different sign.

For this problem, it is not necessary to compute many of the parameters of the fly-by trajectory. Since the planet's velocity with respect to the sun has no component in the $\hat{\mathbf{i}}$ direction, the component of the heliocentric spacecraft velocity in that direction will be identical to that of the vector with respect to Saturn. Therefore, we already know that the final value of the planetocentric velocity vector will be

$$\mathbf{v}_{SC/♄} = 8.627\hat{\mathbf{i}} + 5.495\hat{\mathbf{j}}$$

The heliocentric velocity vector is then

$$\mathbf{v}_{SC/\odot} = \mathbf{v}_{♄/\odot} + \mathbf{v}_{SC/♄} = 8.627\hat{\mathbf{i}} + 15.15\hat{\mathbf{j}}$$

The heliocentric velocity is now $v_{SC/\odot} = 17.43\frac{\text{km}}{\text{s}}$, and the energy is

$$C = \frac{17.43^2}{2} - \frac{\mu_\odot}{a_♄} = \boxed{58.82\frac{\text{km}}{\text{s}}}$$

which is greater than zero. The spacecraft is now on an escape trajectory, out of the solar system. ♠

6.6.2 Fly-bys for Plane Change – The *Ulysses* Mission

One use of gravity assist trajectories is to change the inclination of the heliocentric trajectory. Recall that the impulse needed for plane change is linearly dependent on the velocity; when the velocities are in the tens of kilometers per second, as they are in heliocentric trajectories, this is clearly impossible. However, the use of a fly-by to generate the necessary impulse enables otherwise impossible missions.

An example of this use is the *Ulysses* mission, which put a spacecraft in a nearly polar heliocentric orbit. In order to achieve this orbit, the craft performed a fly-by of Jupiter. The craft was launched from the Shuttle on October 6, 1990, and on February 8, 1992 made its Jupiter fly-by. The resulting orbit had an inclination of 80.2°, and reached perihelion on March 12, 1995[8] Since the fly-by sent Ulysses into the southern hemisphere, the point of fly-by became the descending node, and perihelion is very nearly at the ascending node. The next appearance of the craft at the ascending node was on May 25, 2001, giving an orbital period of about 2260 days.

Such a trajectory is possible because it is possible to choose not just the periapsis of the fly-by hyperbola, but also the position in the planet-centered frame of reference. It takes no more effort to direct the orbit slightly above or below the plane of the ecliptic than it does to change the radial components – recall that we have effectively infinite distance over which to let the effects of a small impulse grow.

Example 6.8. To perform an approximate analysis of the Ulysses fly-by, assume that the craft was at aphelion afterwards. Making the assumption of a circular orbit for Jupiter, this means that $r = 5.2026\ AU$. Taking 2260 days as the period, we have

$$P = 86400 \cdot 2260 = 2\pi a\sqrt{a/\mu_\odot} \quad \Longrightarrow \quad a = 5.042 \times 10^8 \text{ km}.$$

[8] The craft crossed the plane of the ecliptic the next day; the fact that the perihelion is so very close to the ecliptic speaks to the nearly perfect fly-by performed.

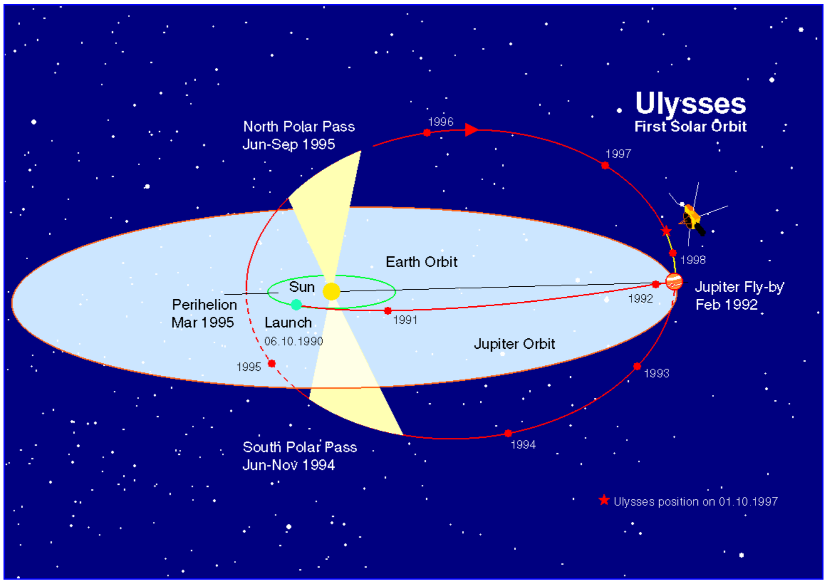


Figure 6.14: Ulysses Earth-to-Jupiter and first solar orbit (from NASA/ESA website)

This provides a perihelion distance of $r_p = 1.54\ AU$, which is incorrect by about 0.15 AU (the actual value is 1.39).

Going with our approximate number, we have that the magnitude of the velocity after the fly-by is $v_a = 8.822\,\mathrm{km \cdot s^{-1}}$. To compute the velocity vector in the local coordinate frame, recall that we have assumed aphelion, so there is no radial component. Therefore

$$\mathbf{v}^+_{SC/\odot} = 0\hat{\imath} + 8.783\cos(80.2°)\hat{\jmath} + 8.783\sin(80.2°)\hat{k} = 1.502\,\hat{\jmath} + 8.693\,\hat{k}.$$

Allowing for the planet's velocity, we have

$$\mathbf{v}^+_{SC/♃} = \mathbf{v}^+_{SC/\odot} - \mathbf{v}_{♃} = -11.557\,\hat{\jmath} + 8.693\,\hat{k}$$

so that the velocity relative to Jupiter is

$$v_\infty = v^+_{SC/♃} = 14.461\,\mathrm{km \cdot s^{-1}}.$$

To proceed, it is necessary to know the velocity vector prior to the gravity assist. This could be found by noting that the transfer time is 490 days, which allows the use of Kepler's equation (assuming that the departure was at perihelion) to compute the true anomaly at Jupiter arriival, and from there the other orbital characteristics. Note that the Hohmann transfer time is 997 days, and the relative velocity at arrival would only be $v^-_{SC/♃} = 5.643\,\mathrm{km \cdot s^{-1}}$. This is much less than that required for the mission. ♠

Note that this example is one situation in which the fast transfer is needed. The final mission objective could not be achieved by executing a Hohmann transfer and a gravity assist.

6.7 A Fairly Involved Example

In this section, the ideas presented in the chapter will be brought together through a single example. While each has been illustrated by examples as they were presented, this may help in seeing how they can be assembled in analyzing a particular mission from end to end.

The problem is a transfer from Earth to Mars. The spacecraft is first launched into a circular Earth orbit of radius 6600 km. The mission planners have determined that the transfer time is not to exceed 200 days. A gravity-assist trajectory around Mars will be determined such that maximizes the heliocentric velocity after the fly-by.

Find the characteristics of the transfer orbit

The first thing to be done is to determine the transfer phase of the mission. To do this, we first make the assumption that the departure from Earth will be in the nature of a Hohmann orbit. That is to say, the velocity of the escape hyperbola will be aligned with that of the Earth as the spacecraft transits from Earth-dominated to interplanetary space. As a result of this, we know that departure occurs at perihelion of the transfer orbit.

As in Example 6.2, determining the transfer orbit characteristics will require an iterative process. In this case, the problem is slightly more complex because we don't yet know the type of transfer orbit. Therefore the first thing to do is to see if the transfer time on a parabolic orbit would be greater or less than the specified time. Recalling Barker's equation (3.31)

$$t - t_p = \frac{h^3}{2\mu^2}\left(\tan\frac{\nu}{2} + \frac{1}{3}\tan^3\frac{\nu}{2}\right)$$

we see that we need both the angular momentum and the true anomaly at arrival to compute the transfer time. Easily enough, we calculate

$$r_p = \frac{p}{1+e} \quad \Longrightarrow \quad p = 2r_p = 2a_\oplus$$

and then apply the relationship $h = \sqrt{\mu p}$ to find that

$$h = \sqrt{2\mu_\odot a_\oplus} = 6.301 \times 10^9 \text{ km}^2/\text{s}$$

on the parabolic trajectory. True anomaly at arrival can then be computed from

$$a_{♂}(1 + \cos\nu) = p \quad \Longrightarrow \quad \nu = 1.2529$$

and the time of flight is found to be $T_{\text{par}} = 69.9$ d. Because the desired time of flight is longer, we conclude that an elliptic transfer orbit will serve.

From this point, Example 6.2 can be used almost directly. Letting a_t denote the semimajor axis of the transfer orbit (we will use a t subscript to denote quantities referring to the transfer orbit throughout this example), we select a value for it. Then, again using $r_p = a_\oplus$, we compute eccentricity from

$$a_\oplus = a_t(1-e)$$

and the eccentric anomaly then from

$$a_{\mars} = a_t(1 - e\cos E)$$

Finally,

$$T_t = \sqrt{a_t^3/\mu_\odot}\,(E - e\sin E)$$

provides the transfer time, and we correct the estimate of a_t until this equals 200 days.

Noting that the Hohmann transfer time is the longest of all two-impulse transfers, the associated semimajor axis is a lower bound. Trying $a_t = 2a_H$ produces a time of flight of only 93.07 d. Using bisection produces the following table

a_t/a_H	e	E	T_t (days)
2.0000	0.6038	0.8548957	93.075662
1.5000	0.4716752	1.1446092	108.25451
1.2500	0.3660102	1.4777967	128.21249
1.1250	0.2955669	1.8215656	150.94981
1.0625	0.2541297	2.137743	173.57354
1.03125	0.2315275	2.4012447	193.73236
1.015625	0.2197049	2.6059719	210.32709
1.0234375	0.2256613	2.4931648	201.07405
1.0273438	0.2286056	2.445175	197.21978
1.0253906	0.2271363	2.4686073	199.09583
1.0244141	0.2263995	2.4807374	200.0714

Several more iterations could be done to get an even closer answer. However, at some point the accuracy of the calculations becomes pointless in light of the simplifications made in the Patched Conic approximation. Achieving eight digits of agreement to a model that is only accurate to a few per cent is pointless. We thus take the characteristics of the transfer orbit to be

$$a_t = 1.9338 \times 10^8\text{km} \qquad \text{and} \qquad e = 0.2264$$

It is interesting to note that an increase in the semimajor axis of only two and a half per cent has decreased the transfer time by nearly 60 days relative to the Hohmann transfer. Looking farther up the table shows that a reduction of 49 more days (to 150.9 days) would require a further ten per cent increase.

Compute the necessary departure impulse

Having found the transfer orbit, it is now possible to compute the impulse needed to transition from the parking orbit to the escape hyperbola.[9] Again, we have already seen a very similar problem, in this case in Example 6.4. Because the velocity vector as the spacecraft leaves Earth's sphere of influence (that is, as $r \longrightarrow \infty$ on the departure hyperbola) is aligned with that of the Earth, the arithmetic can be done without recourse to vectors. From the results of the last section, we compute the velocity at perihelion of the transfer orbit, and subtract the velocity of the Earth to get the excess speed of the departure hyperbola.

$$\begin{aligned} C_t &= -\mu_{\odot}/2a_t = -343.15\ \mathrm{km}^2/\mathrm{s}^2 \\ v_p &= \sqrt{2\,(C_t + \mu_{\odot}/a_{\oplus})} = 32.985\,\mathrm{km/s} \\ &\Longrightarrow \quad v_\infty = v_p - \sqrt{\mu_{\odot}/a_{\oplus}} = 32.985 - 29.785 = 3.200\ \mathrm{km/s}. \end{aligned}$$

The velocity on the parking orbit is found from the specified radius $a_1 = 6600$ km to be

$$v_c = \sqrt{\mu_{\oplus}/a_1} = 7.771\ \mathrm{km/s}.$$

The velocity on the departure hyperbola at $r = a_1$ is computed from the energy relation as

$$v = \sqrt{2\,(C_h + \mu_{\oplus}/a_1)} = \sqrt{v_\infty^2 + 2\mu_{\oplus}/a_1} = 11.447\ \mathrm{km/s}.$$

The necessary impulse is then simply

$$\Delta v = v - v_c = 3.675\ \mathrm{km/s}.$$

Find the heliocentric velocity at arrival

Following Example 6.5, we compute the magnitude of the arrival velocity, and then the individual components. The magnitude can be computed as usual from the energy equation,

$$v_{SC/\odot} = \sqrt{2\left(C_t + \frac{\mu_{\odot}}{a_{♉}}\right)} = 21.867\ \mathrm{km/s}.$$

The components may then be calculated from the angular momentum. As in Example 6.5, recall that

$$h = rv\sin\phi$$

at any point on any trajectory. At departure, this is just

$$h_t = r_p v_p = a_{\oplus} v_p = 4.934 \times 10^9\ \mathrm{km}^2/\mathrm{s}$$

[9] We have not discussed launch directly from Earth surface into escape trajectory. This is possible, and in fact is often done, but the details of launch through the atmosphere make the problem beyond the scope of this chapter.

(recall that v_p was computed above). At arrival, the value of r is simply the orbital radius of Mars, so

$$\phi = \sin^{-1}\left(\frac{h_t}{a_{♂} v_{SC/\odot}}\right) = 1.4291 \ \ (81.88°)$$

The heliocentric velocity is then

$$\mathbf{v}_{SC/\odot} = v_{SC/\odot}\left[\cos\phi\hat{\imath} + \sin\phi\hat{\jmath}\right] = 3.088\hat{\imath} + 21.65\hat{\jmath} \text{ km/s}. \tag{6.18}$$

Considering (6.18), note that the component of the velocity in the $\hat{\jmath}$ direction is

$$\mathbf{v}_{SC/\odot} \cdot \hat{\jmath} = v_{SC/\odot} \sin\phi.$$

If we do not need the included angle for other purposes, we could then compute the components directly from

$$\mathbf{v}_{SC/\odot} \cdot \hat{\jmath} = h_t/a_{♂}$$

and

$$(\mathbf{v}_{SC/\odot} \cdot \hat{\imath})^2 = v^2_{SC/\odot} - (\mathbf{v}_{SC/\odot} \cdot \hat{\jmath})^2 = v^2_{SC/\odot} - (h_t/a_{♂})^2.$$

Compute the Mars-relative arrival velocity

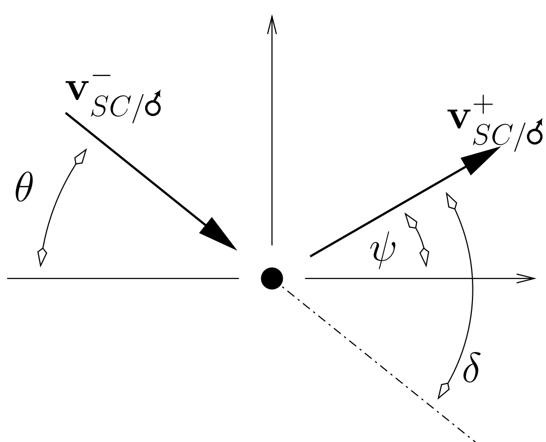


Figure 6.15: Martian fly-by.

Continuing to follow 6.5, this is simply the heliocentric velocity of the spacecraft, minus the velocity of the planet. The radial component remains unchanged, and

$$\begin{aligned}
\mathbf{v}_{SC/♂} &= \mathbf{v}_{SC/\odot} - \mathbf{v}_{♂} \\
&= 3.088\hat{\imath} + \left(21.65 - \sqrt{\mu_{\odot}/a_{♂}}\right)\hat{\jmath} \\
&= 3.088\hat{\imath} - 2.482\hat{\jmath} \text{ km/s}.
\end{aligned}$$

The magnitude of the vector is the excess velocity of the arrival hyperbola (in this case, the fly-by hyperbola, since we will not be staying):

$$v_\infty = \|\mathbf{v}_{SC/♂}\| = 3.962 \, \text{km/s}.$$

Find the gravity-assist trajectory that maximizes heliocentric velocity

Recalling the patch conditions and especially Figure 6.12, it is clear that the maximum velocity after the fly-by will be generated by most closely aligning the departure leg of the fly-by hyperbola with the velocity of Mars. This particular case is illustrated in Figure 6.15. It is obvious that the post fly-by velocity will be maximized if the turning angle is maximized (assuming it does not thereby rotate the velocity vector beyond the y-axis, which is unlikely), and that the vector should be rotated in the counter-clockwise direction.

It can easily be shown that the turning angle is maximized when the periapsis of the fly-by hyperbola is minimized. The radius of the planet is an obvious lower limit, so setting $r_p = r_{\male} = 3393\,\mathrm{km}$, we have from (6.17)

$$e = \frac{r_p v_\infty^2}{\mu} + 1 = \frac{r_{\male} v_{SC/\male}^2}{\mu_{\male}} + 1 = 2.2372$$

leading to a turning angle of

$$\delta = 2\sin^{-1}\left(\frac{1}{e}\right) = 0.9268 \;\; (53.10^\circ).$$

From the sketch, we see that

$$\mathbf{v}_{SC/\male}^{+} = v_\infty \left(\cos\psi\hat{\imath} + \sin\psi\hat{\jmath}\right)$$

and plugging in $\psi = \delta - \theta$, and noting that

$$\theta = \tan^{-1}\left(\frac{2.482}{3.088}\right) = 0.6769 \;\; (38.78^\circ),$$

we have for the post-fly-by velocity

$$\mathbf{v}_{SC/\male}^{+} = 3.839\hat{\imath} + 0.980\hat{\jmath}\,\frac{\mathrm{km}}{\mathrm{s}}.$$

Adding the velocity of the planet, we have the heliocentric velocity now to be

$$\mathbf{v}_{SC/\male}^{+} = 3.839\hat{\imath} + 25.11\hat{\jmath}\,\frac{\mathrm{km}}{\mathrm{s}},$$

giving a magnitude of $v = 25.40$ km/s. This is about 3.53 km/s greater than before the fly-by, though still well below the necessary velocity to escape from the solar system.

6.8 Planetary Ephemeris

Throughout this chapter, we have made extensive use of the approximation that the planetary orbits are circular. While this is certainly a good first approximation, it is of course not completely true. However, once the inclination and eccentricity of the orbits are brought into the problem, solutions require the use of computer iteration.

Sufficient information for the use of patched conics for elementary analysis is included in the tables in Appendix E.3. However, those tables list only the mean characteristics of the orbits themselves, with no data from which to compute the position of the planets on the orbits. Further, because the solar system is much more complex than a two-body system of the type we have investigated in this book, the elements of the planetary orbits are not constant. The topic of variation of elements will be covered in detail in a later chapter. Here, we mention only that for accurate computation of positions, these variations must be taken into account. A table of planetary positions, or algorithms and constants from which position can be computed, is known as an *ephemeris*.

Whether the positions are tabulated as functions of time, or the orbital elements are expressed as functions of time, there is a need to better define the terms "time" and "date" themselves.

6.8.1 Julian Dates and Ephemeris Time

The everyday definition of time is the division of the solar day into hours, minutes, and seconds. This is perfectly good for most purposes of most people, but not quite sufficient for astrodynamics. For most people as well, the calendar date is enough to know. However, astrodynamics deals with time spans of centuries, and over the centuries the calendars have changed. It is not enough to state that an event happened on a particular date; it must be known on which calendar the date is specified. To make matters even more confusing, various calendars were adopted at different times in different parts of the world.

For astrodynamical work, the date is specified as the *Julian day number*. This is the number of days since January 1, 4713 BC. The Julian day begins at noon Greenwich mean time, so that for example midnight on January 1, 1999 (or 1999 Jan. 1.0) is JD 2451179.5.

There are a few different algorithms for finding the Julian day number for a given calendar date. Meeus ([10], chapter 7) gives the following for a date in the Gregorian calendar. Letting y be the year, m the month, and d the day (including decimal), proceed as

$$Y = \begin{cases} y & \text{if} \quad m > 2 \\ y-1 & \text{if} \quad m \leq 2 \end{cases} \tag{6.19}$$

$$M = \begin{cases} m & \text{if} \quad m > 2 \\ m+12 & \text{if} \quad m \leq 2 \end{cases} \tag{6.20}$$

$$A = \texttt{floor}\,(Y/100) \tag{6.21}$$

$$B = 2 + \texttt{floor}\,(A/4) - A \tag{6.22}$$

$$\begin{aligned} \mathrm{JD} = {} & \texttt{floor}(365.25(Y+4716)) \\ & + \texttt{floor}(30.6001(M+1)) + d + B - 1524.5 \end{aligned} \tag{6.23}$$

The function `floor` above means to round towards negative infinity;

$$\texttt{floor}(7.5) = 7; \qquad \texttt{floor}(-7.5) = -8.$$

If the calendar date is already in the Julian calendar, the term B in (6.22) above is taken to be zero. The Julian calendar was in use before the Gregorian, which was introduced in 1582. However, the calendar was not adopted uniformly; Great Britain had not fully adopted it until 1752, for example, and some countries did not accept it until the early 20$^{\text{th}}$ century.

For dates from March 1, 1900 through February 28, 2100, $B = -13$. Some single-line formulae are available for dates within these limits. However, the method outlined above is valid for all dates after 4712 BC, so long as the Julian-Gregorian calendar distinction is kept in mind.

Example 6.9. Find the Julian day number for 9:00 a.m. GMT on August 8, 1971.

August is the eighth month, so $m = 8$, and of course $y = 1971$. The day number is $d = 8.375$ (eight plus 9/24 days). Then

$$\begin{aligned}
Y &= 1971 \\
M &= 8 \\
A &= \texttt{floor}(1971/100) \;=\; 19 \\
B &= 2 + \texttt{floor}(19/4) - 19 \;=\; 2 + 4 - 19 \;=\; -13 \\
\text{JD} &= \texttt{floor}(365.25 \cdot 6687) + \texttt{floor}(30.6001 \cdot 9) + 8.375 - 13 - 1524.5 \\
&= 2441171.875
\end{aligned}$$

Note that the time of day (GMT), expressed as the fractional part of the day number, has shifted by 0.5 as a result of being measured from noon of the previous day, rather than midnight. ♠

Long periods of time are measured in *Julian centuries*, each of which is exactly 36525 mean solar days long.

The matter of *time* is also more complicated than might be expected. Most people think of a "day" as the time between one noon and the next; "noon" in turn is thought of as the time during the day at which the sun is highest in the sky. This is good enough for most things,[10] but not for astrodynamics.

The first consideration is that the *solar* day is longer than the *sidereal* day. The term sidereal means "with respect to the stars", and a sidereal day is length of time between successive passages of a distant star over a particular longitude of the earth. The *solar* day is the usual day just discussed above. From Figure 6.16, it can be seen that the earth must rotate only 360°for the first, but 360°plus its angular motion about to sun for the second (this is $\Delta\theta$ in the figure).

[10] Including, almost always, government work

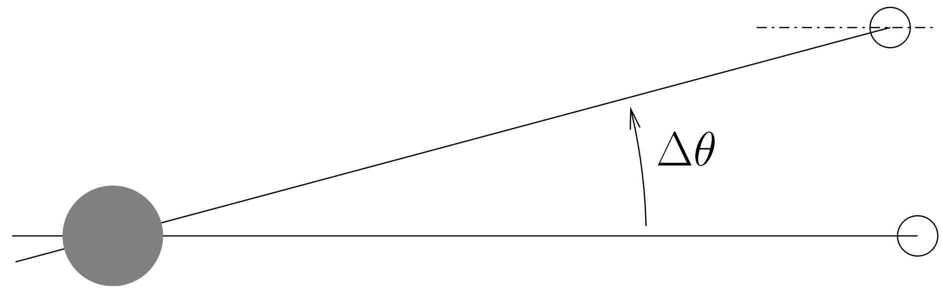


Figure 6.16: Sidereal vs. solar motion of a planet.

Even allowing for this, the time from noon to noon is not constant. This is partly because the earth's orbit is elliptic, so that it moves less per day near aphelion than at perihelion. Also, the earth's rotation slows due to tidal friction. Finally, it has been discovered that the rotation is also subject to sudden slowdowns, for reasons not yet understood.

For all these reasons, the astronomical community uses *ephemeris time*, which is a fictional time tied to a perfectly predictable earth. That is to say, ephemeris time is the independent variable in the astronomers' models of celestial mechanics, and Greenwich mean Time is the time that people actually use, subject to all the unknown and unpredictable perturbations. Corrections are published after long series of observations.

These very fine points of time-keeping are beyond our needs for the purposes of this book. Their existence is instructive, but for us, mean solar or sidereal time will be sufficient. By definition, the *mean solar day* is 86400 *mean solar seconds*, and this is very nearly the average time from noon to noon across the year ("nearly" because, as mentioned above, this value changes). The *sidereal day* is close enough to 86164 mean solar seconds long.

6.8.2 Mean Planetary Elements

Because the solar system is much more complex than a two-body system of the type we have investigated in this book, the elements of the planetary orbits are not constant. The topic of variation of elements will be covered in detail in a later chapter. Here, we mention only that for accurate computation of positions, these variations must be taken into account. There are several methods for doing so. The most accurate is to use a highly accurate model of the solar system, and start with the best observed positions available at a particular time. From these initial conditions, and using a very good numerical integration scheme, the positions and velocities of the planets can be computed well into the future. Naturally, as mentioned in chapter 4, the motion of the axes themselves must be dealt with, as well.

Somewhat less accurate is to express the elements as functions of time. Since there is no explicit function[11] for the planetary elements, this approach requires that the function be approximated in some way. The theory of such approximations is outside the scope of this text. It will suffice to say that the existing expressions have many terms, and that the greater the desired accuracy of the result, the more terms are required in the computation.

One factor that complicates the matter is that there are both secular and periodic variations in the elements of the planetary orbits. Highly accurate formulae will include both of these terms; Meeus devotes 40 pages to "the most important periodic terms"[12] The periodic variations are not, however, very large, and for our purposes we will consider only the *mean* values of the elements. These are approximate values of the elements of the orbit associated with the planetary position and velocity at a particular time.

These values can be expressed to a sufficient degree of accuracy by low-order polynomials in time. "Low precision" is more than sufficient to assess propulsion needs and perform most mission analysis. Various authors present linear, quadratic, and cubic approximations. The independent variable is usually Julian centuries, and for any term beyond the linear correction to appear in the first several decimal places generally requires at least ten to 100 centuries. Thus, we will content ourselves with a linear expression, as

$$f(T) = a_0 + a_1 T.$$

Table 6.1 comes from the Solar System Dynamics Group at the Jet Propulsion Laboratory.[13] gives the constant term and linear corrections for the nine planets. These are referenced to the heliocentric axis frame on the epoch J2000.0 (Julian day JD2451545.0). For each planet, the first row gives the constant terms, and the second the corresponding centennial rates. The semimajor axis is given in astronomical units, and the angular elements in degrees. The rates of the angular elements are given as degrees per century. The table was computed to best fit each element over the period 1800 to 2050 AD. Outside of this 250 year span, the accuracy is questionable, but for our purposes will suffice.

Note that the elements given are the usual for semimajor axis, eccentricity, inclination, and right ascension of the ascending node. The last two are the longitude of perihelion, which we recall is defined as $\varpi = \Omega + \omega$, and the mean longitude. These were mentioned in section 4.6. The mean longitude is defined as

$$L = \varpi + M.$$

[11] In this case, we mean that there is no function that can be expressed as a finite series of simple polynomial or trigonometric terms, and that is completely accurate

[12] [10], page 413.

[13] "Keplerian Elements for Approximate Positions of the Major Planets", E.M. Standish, Solar System Dynamics Group, JPL/Caltech, Pasadena, California. Online at `http://ssd.jpl.nasa.gov/?planet_pos`.

	a	e	ι	L	ϖ	Ω
☿	0.38709927	0.20563593	7.00497902	252.25032350	77.45779628	48.33076593
	0.00000037	0.00001906	-0.00594749	149472.67411175	0.16047689	-0.12534081
♀	0.72333566	0.00677672	3.39467605	181.97909950	131.60246718	76.67984255
	0.00000390	-0.00004107	-0.00078890	58517.81538729	0.00268329	-0.27769418
⊕	1.00000261	0.01671123	-0.00001531	100.46457166	102.93768193	0.0
	0.00000562	-0.00004392	-0.01294668	35999.37244981	0.32327364	0.0
♂	1.52371034	0.09339410	1.84969142	-4.55343205	-23.94362959	49.55953891
	0.00001847	0.00007882	-0.00813131	19140.30268499	0.44441088	-0.29257343
♃	5.20288700	0.04838624	1.30439695	34.39644051	14.72847983	100.47390909
	-0.00011607	-0.00013253	-0.00183714	3034.74612775	0.21252668	0.20469106
♄	9.53667594	0.05386179	2.48599187	49.95424423	92.59887831	113.66242448
	-0.00125060	-0.00050991	0.00193609	1222.49362201	-0.41897216	-0.28867794
⛢	19.18916464	0.04725744	0.77263783	313.23810451	170.95427630	74.01692503
	-0.00196176	-0.00004397	-0.00242939	428.48202785	0.40805281	0.04240589
♆	30.06992276	0.00859048	1.77004347	-55.12002969	44.96476227	131.78422574
	0.00026291	0.00005105	0.00035372	218.45945325	-0.32241464	-0.00508664
♇	39.48211675	0.24882730	17.14001206	238.92903833	224.06891629	110.30393684
	-0.00031596	0.00005170	0.00004818	145.20780515	-0.04062942	-0.01183482

Table 6.1: Best-fit planetary elements and centennial rates.

Were the planetary orbits perfectly keplerian, the rate of the mean longitude would simply be the mean motion.

It should also be noted that the values for the earth are actually those for the center of mass of the earth-moon pairing.

Example 6.10. What were the orbital elements of Jupiter on February 8, 1992, when the Ulysses spacecraft made its fly-by?

The Julian day number for February 8, 1992 is 2448660.5. This was

$$\frac{2448660.5 - 2451545.0}{36525} = -0.078973$$

Julian centuries before epoch for the data in the table. Compute the semimajor axis, eccentricity, inclination, and right ascension of the ascending node directly from the table as

$$\begin{aligned}
a &= 5.20288700 - 0.00011607(-0.078973) = 2.20290\ AU \\
e &= 0.04838624 - 0.00013253(-0.078973) = 0.04840 \\
\iota &= 1.30439695 - 0.00183714(-0.078973) = 1.30454^\circ \\
\Omega &= 100.47390909 + 0.20469106(-0.078973) = 100.458^\circ
\end{aligned}$$

The argument of periapsis is computed from the definition of longitude of periapsis as

$$\begin{aligned}
\omega &= \varpi - \Omega \\
&= 14.72847983 + 0.21252668(-0.078973) - 100.458 \\
&= -85.7460^\circ
\end{aligned}$$

The true anomaly is found by first computing the mean longitude:

$$L = 34.39644051 + 3034.74612775(-0.078973) = -205.267,$$

and then the mean anomaly as

$$M = L - \varpi = -205.267 - 14.7119 = -219.978$$

which modulo 360 is $M = 140.022^\circ$. Iteratively solving Kepler's equation for eccentric anomaly gives $E = 141.739^\circ$. Finally, the true anomaly is found from (3.5) to be $\nu = 143.424^\circ$. ♠

For purposes of mission planning, the true longitude is also useful. Time of flight calculations depend on knowing the angle swept by the craft during the transfer. While this can be calculated from the combination of true anomaly, right ascension, and argument of periapsis, these are combined conveniently in the true longitude.

6.9 Problems

1. Consider a mission from the Earth to Venus and back, assuming Hohmann transfer each way. Compute

   (a) The time of flight in each direction.

   (b) The synodic period.

   (c) The necessary waiting period before return, and the overall mission duration.

2. A spacecraft is in a circular parking orbit of $a = 7000\,\mathrm{km}$ about the Earth. It is to be sent on a *parabolic* transfer trajectory to Mars, and will be at perihelion at Earth escape.

   (a) What is the necessary hyperbolic escape speed for the departure hyperbola from the Earth?

   (b) What is the impulse needed for injection into the escape orbit?

   (c) What is the time of flight to Mars on the heliocentric parabola?

   (d) What is the velocity of the craft when it reaches Mars orbit?

   (e) What is the necessary phase angle at departure to achieve rendezvous with Mars?

3. Consider a mission from the Earth to Mercury. After a Hohmann transfer, the vehicle will pass the planet on the far side from the Sun. The closest approach to the planet will be an altitude of 61 km ($r_p = 2500\,\mathrm{km}$). Find:

   (a) The velocity offset required to achieve the specified periapsis radius.

   (b) The eccentricity and turning angle of the fly-by hyperbola.

   (c) The semi-major axis and eccentricity of the heliocentric orbit resulting from the fly-by.

   (d) The closest approach the vehicle will have to the Sun, and the time of flight after fly-by to perihelion.

   Note that Mercury is quite small relative to the high velocity the craft will have after the transfer.

4. We have considered departure from a circular parking orbit by a single-impulse transition to the departure hyperbola. Hermann Oberth, however, proposes a two-impulse solution as in Figure 6.17.

   Suppose that we desire an excess velocity of $v_\infty = 2.95\,\mathrm{km \cdot s^{-1}}$ from the earth. Starting from a circular orbit with $a_{\mathrm{park}} = 3.84 \times 10^5\,\mathrm{km}$,

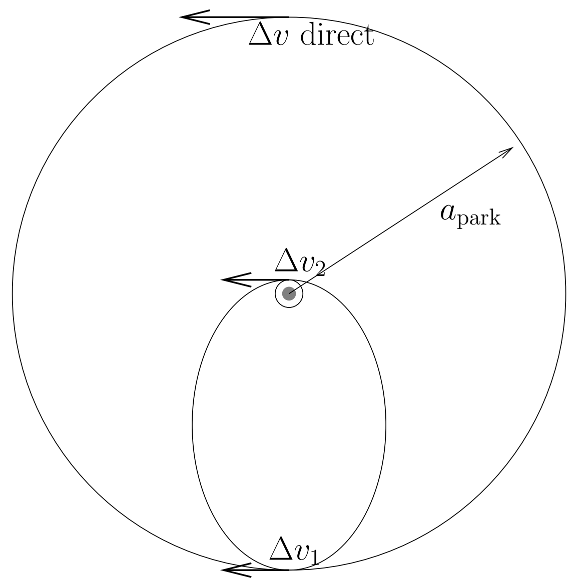


Figure 6.17: Oberth two-impulse escape.

(a) Compute the required direct Δv for the single-impulse approach.

(b) Consider the two-impulse approach. Use the first impulse to put the craft on an ellipse with a perigee of $r_p = 6578$ km. The second will transition the craft from perigee of the ellipse to perigee of an escape hyperbola. What is the total Δv required for this approach.

The circular Lagrange points of the Earth-Moon system are at $a = 3.84 \times 10^5$ km (this will be studied in a later chapter).

5. Consider an earth to Jupiter transfer that arrives with $\nu = 165°$ on the transfer orbit.

   (a) What is the transfer time?

   (b) What is the velocity at arrival relative to Jupiter?

   (c) Compute a fly-by with a turning angle of 90°(that is, find the eccentricity and periapsis radius). If the fly-by is used to slow down the craft and turn it towards the Sun, what are the eccentricity, semimajor axis, and perihelion of the resulting trajectory?

   Use the circular orbit assumptions in doing this problem.

6. Consider the Ulysses mission as in example 6.8. Supposing a Hohmann transfer to Jupiter, what is the maximum inclination that could be achieved for a post-encounter heliocentric trajectory? What would the eccentricity and perihelion

distance of the resulting trajectory be? (Note that the maximum inclination will be achieved for a post-encounter trajectory that has no radial velocity component.) Use the circular orbit assumptions in this problem.

7. Consider the example of section 6.7. Assume the fly-by hyperbola skims the planetary surface, and ignore any effects due to the Martian atmosphere. What would be the effect on the final heliocentric velocity of a 1 km/s impulse applied at periapsis of the fly-by hyperbola? Assume the post fly-by orbit remains in the plane of the ecliptic, and assume that the orbit of Mars is circular.

8. Again, consider the Ulysses mission from section 6.6.2.

    (a) Using the information from table 6.1, find the true longitude of the Earth at launch of the Ulysses mission (recall the definition of true longitude from 4.6), and of Jupiter at arrival.

    (b) Ignoring inclination, take the difference in true longitude as the angle swept during flight. Assuming launch at perihelion, find the characteristics of the transfer orbit that achieves Jupiter arrival in 490 days.

    (c) What is the resulting velocity vector relative to Jupiter at arrival? Include the effects of eccentricity of Jupiter's orbit (again, neglect inclination).

    (d) From the results of part 8c, can you create a fly-by that achieves an 80 degree inclination heliocentric orbit? Give the details of the fly-by hyperbola (r_p and eccentricity), and the resulting helioentric orbit.

    Note that ignoring inclination will cause some discrepancy between the results and the actual mission.

9. Among the most famous and dramatic space probe missions are the grand tours of Voyager One and, especially, Voyager Two. Their heliocentric trajectories are shown in Figure 6.18. For a basic analysis, ignore the effects of inclination in this problem.

    (a) Find the Julian Day numbers for the Voyager Two encounters of Jupiter, Saturn, and Uranus.

    (b) Find the true longitudes of each planet, and their distances from the sun, on the corresponding days.

    (c) Find the characteristics of the transfer orbits from Jupiter to Saturn, and from Saturn to Uranus.

    (d) Find an approximate fly-by trajectory about Saturn. Because ignoring the inclination causes small errors, an exact solution will not be possible. Instead, find the fly-by that achieves the correct turning angle. What is the error in the resulting heliocentric velocity?

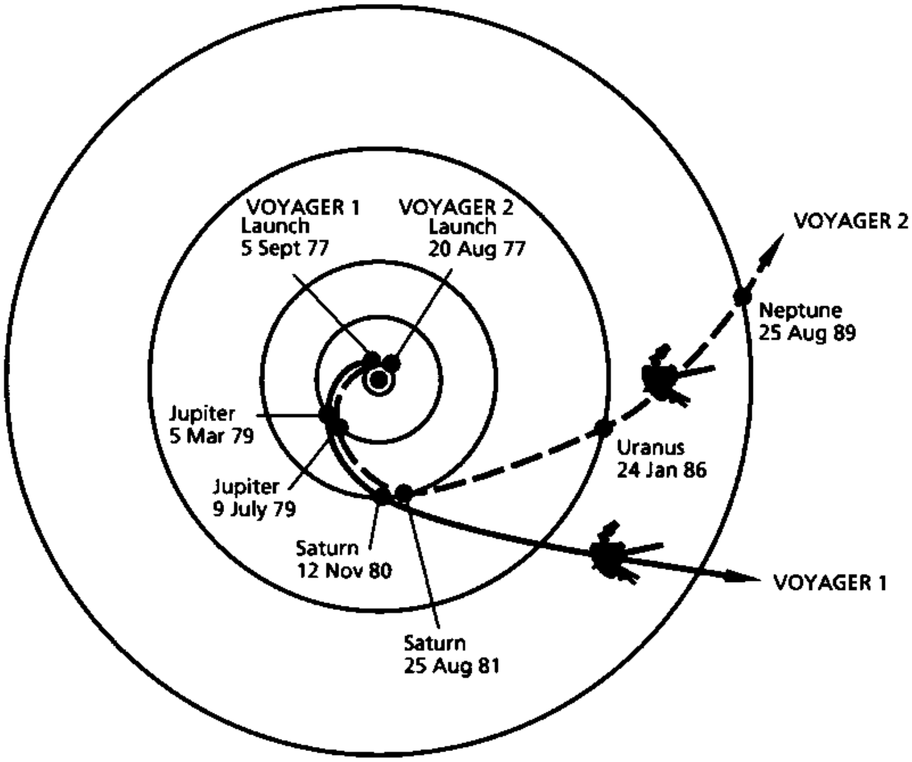


Figure 6.18: Grand tours of Voyagers 1 and 2.

Note that to do this part of the problem it will be necessary to compute the Saturn-relative velocity of the craft before and after fly-by. Include the effect of eccentricity in computing the velocity of Saturn.

The approach of this problem makes the simplification that the planetary encounters take place at essentially a single point from the heliocentric point of view, as well as ignoring inclination. The resulting errors make it impossible to exactly match the actual mission. However, the results are sufficiently close to show that the methods laid out in this chapter provide an excellent first approximation.

10. Consider a preliminary design of Voyager One, as in Figure 6.18. In this problem, we will not attempt to match the actual voyage.

    (a) Assume that the launch date has been established as September 5, 1977. On that date, the true longitude of the earth was 342.63°, and that of Jupiter was 62.23°. Using the usual patched conic approximations, compute a trajectory from the Earth that begins at perihelion, and will achieve Jupiter rendezvous.

(b) The true longitude of Saturn at launch was 141.59°. Assuming the dates of encounter in Figure 6.18 are correct, find the semimajor axis and eccentricity of the transfer orbit from Jupiter to Saturn.

Chapter 7

Restricted Three-body Problem

The term "three body problem" is most correctly used for any problem involving three masses and their mutual attraction. The *restricted* three-body problem is much more specialized, and was first formulated by Leonard Euler in 1772.

7.1 Formulation

The problem is usually formulated in canonical form as follows: There are two primary bodies, the smaller of mass μ and the larger of mass $1-\mu$. The distance between the bodies is taken to be unity, and is assumed constant. The time unit is defined such that one orbit of the two-body system requires 2π units. In this set of units, the rotational rate is 1 and the value of the gravitational constant G comes out to be unity, as well.

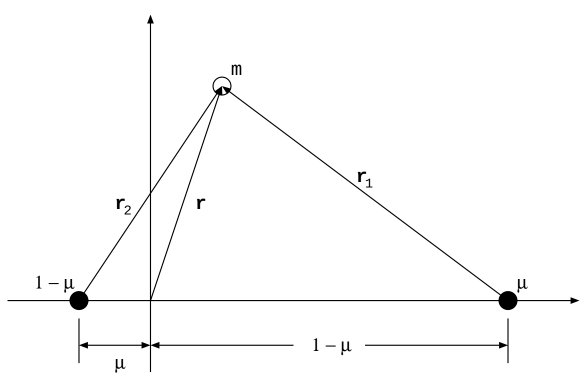


Figure 7.1: Geometry of three-body problem.

To this system of two bodies we add a third. The third body has very small mass, so that it has no measurable effect on the motion of the primaries (such a situation occurs in the instance of a spacecraft in the vicinity of the Earth and Moon, for example).

As the distance between the primaries is constant, it must be that they describe simple circular orbits about each other and, therefore, about the center of mass. We will describe the motion of the infinitesimal mass m with respect to the rotating frame in which $\hat{i}$ lies on the line running through the centers of the two primaries, $\hat{j}$ is in the orbital plane, and $\hat{k}$ completes the right-handed system. This frame has a rotational rate with respect to the inertial frame of $\omega = 1\hat{k}$.

The equations of motion of the system are derived from basic mechanics. Since they will be derived in a coordinate frame that is not inertial, we must allow for the rotation of the local frame. We have

$$\ddot{\mathbf{r}} = \frac{\partial^2 \mathbf{r}}{\partial t^2} + 2\omega \times \frac{\partial \mathbf{r}}{\partial t} + \omega \times \omega \times \mathbf{r} = \frac{\mathbf{F}}{m}$$

where the partial derivatives are taken with respect to the rotating frame, and $\mathbf{r} = x\hat{i} + y\hat{j} + z\hat{k}$ is the vector from the center of mass to the mass m.

The gravitational attraction is found from Newton's law to be

$$\mathbf{F} = \frac{-\mu G m \mathbf{r}_1}{r_1^3} - \frac{(1-\mu) G m \mathbf{r}_2}{r_2^3}$$

thus we have

$$(\ddot{x}\hat{i} + \ddot{y}\hat{j} + \ddot{z}\hat{k}) + 2\hat{k} \times (\dot{x}\hat{i} + \dot{y}\hat{j} + \dot{z}\hat{k}) + \hat{k} \times \hat{k} \times (x\hat{i} + y\hat{j} + z\hat{k}) = \frac{-\mu \mathbf{r}_1}{r_1^3} - \frac{(1-\mu)\mathbf{r}_2}{r_2^3}$$

(recall $\omega = 1$ and $G = 1$).

Now, $\mathbf{r}_1 = \mathbf{r} - (1-\mu)\hat{i}$ and $\mathbf{r}_2 = \mathbf{r} + \mu\hat{i}$. Working out the cross-products, substituting for $\mathbf{r}_1$ and $\mathbf{r}_2$, and equating the components gives

$$\ddot{x} - 2\dot{y} - x = \frac{-(x-1+\mu)\mu}{r_1^3} - \frac{(x+\mu)(1-\mu)}{r_2^3} \tag{7.1}$$

$$\ddot{y} + 2\dot{x} - y = \frac{-\mu y}{r_1^3} - \frac{(1-\mu)y}{r_2^3} \tag{7.2}$$

$$\ddot{z} = \frac{-\mu z}{r_1^3} - \frac{(1-\mu)z}{r_2^3} \tag{7.3}$$

What we have derived is a highly non-linear set of of coupled second-order equations (the coupling between the $\ddot{z}$ equation and the other two is hidden in r_1 and r_2, which are functions of x, y, and z). They were first derived over 200 years ago, and it has since been proven that there is no general closed-form solution. We can, however, derive a great deal of information about the system.

7.2 Equilibrium Points

If we can't solve the equations of motion, we can at least look for equilibrium points. An equilibrium point in a dynamical system is one at which all of the velocities and accelerations are zero.

The definition of an equilibrium point makes looking for one straight-forward. We simply set all the velocities and accelerations to zero and solve the resulting equations for x, y, and z.

Consider first the equation for $\ddot{z}$. Setting the right hand side to zero gives

$$\frac{\mu z}{r_1^3} = -\frac{(1-\mu)z}{r_2^3}$$

For this to be true, either $z = 0$ or

$$\left(\frac{r_2}{r_1}\right)^3 = \frac{\mu - 1}{\mu} \tag{7.4}$$

Since r_1 and r_2 are magnitudes of vectors, they are clearly positive, and since $0 < \mu < 1$, the right hand side of (7.4) is obviously negative. Thus, the only possible solution is $z = 0$, giving us that all equilibrium points lie in the xy plane.

Consider the other two equations. Setting $\ddot{x} = \dot{x} = \ddot{y} = \dot{y} = 0$, we get

$$x = \frac{(x-1+\mu)\mu}{r_1^3} + \frac{(x+\mu)(1-\mu)}{r_2^3} \tag{7.5}$$

$$y = \frac{-\mu y}{r_1^3} + \frac{(1-\mu)y}{r_2^3} \tag{7.6}$$

With z set to zero, the expressions for r_1 and r_2 become

$$\begin{aligned} r_1 &= \sqrt{(x-1+\mu)^2 + y^2} \\ r_2 &= \sqrt{(x+\mu)^2 + y^2} \end{aligned}$$

Euler made the further simplification of assuming $y = 0$, and looking for points along the x axis. With this assumption, (7.6) reduces to an identity, and (7.5) becomes

$$x = \frac{(x-1+\mu)\mu}{\|x-1+\mu\|^3} + \frac{(x+\mu)(1-\mu)}{\|x+\mu\|^3}$$

We can clear the denominators and arrive at a quintic equation for x. Unfortunately, quintics cannot be solved analytically in general. However, for $0 \leq \mu \leq 1$, it turns out that the polynomial has only three real roots. It can be shown that the solution always reduces to the situation shown in figure Figure 7.2 below.

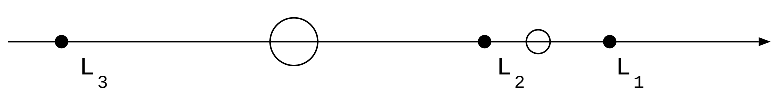


Figure 7.2: X-axis Lagrange points.

Of these three roots, two bracket the smaller primary, while the third lies at *about* unit distance (depending on the value of μ) on the other side of the larger primary.

There are two other points of equilibrium, discovered by Lagrange, also in 1772. In his honor, all of the equilibrium points are denoted by L_i, and generally referred to as Lagrange points (they are also sometimes known as *libration* points, though the term "libration" has a different meaning in astronomy). To find these, return to our equilibrium equations (7.5) and (7.6). Now suppose $r_1 = r_2 = 1$. Then the equations reduce to

$$x = (x - 1 + \mu)\mu + (x + \mu)(1 - \mu) \tag{7.7}$$

$$y = \mu y + (1 - \mu)y \tag{7.8}$$

Multiplying these equations out shows them to be identities – they are automatically satisfied. This means that the two points shown in Figure 7.3, sometimes referred to as the *triangular* Lagrange points, are also equilibrium points.

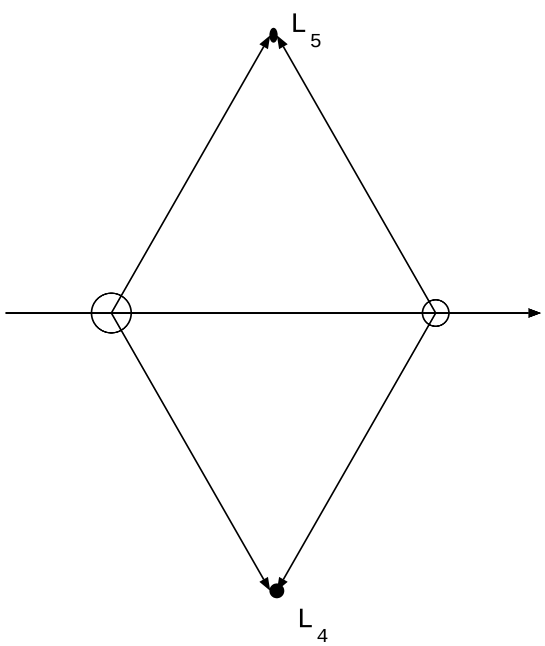


Figure 7.3: Triangular Lagrange points

Since the orbital radius is defined to be unity in the problem formulation, this means that the triangular Lagrange points are the same distance from each primary as the primaries are from each other. In practice, μ is generally very small and the smaller primary is thought of as orbiting the larger. This means that the Lagrange points are in the orbit of the smaller primary, leading and following by sixty degrees of arc.

There are at least two instances in our solar system in which the Lagrange points are important. The equilibrium points in the Earth-Moon system are naturally of interest. The other set is the Lagrange points of the Sun-Jupiter system. In this case, the L_4 and L_5 points are populated by the Trojan asteroids.

7.3 Stability of the Lagrange Points

We will use the results of appendix D to analyze the stability of motion near the Lagrange points of our canonical system.

Defining as in section D.1

$$x = \tilde{x} + \xi \qquad y = \tilde{y} + \eta \tag{7.9}$$

where $(\tilde{x}, \tilde{y})$ is a Lagrange point, we take derivatives and establish the relationships

$$\dot{x} = \dot{\xi} \qquad \dot{y} = \dot{\eta} \tag{7.10}$$

$$\ddot{x} = \ddot{\xi} \qquad \ddot{y} = \ddot{\eta} \tag{7.11}$$

We use Taylor's theorem to expand the equations of motion (7.1,7.2,7.3) as

$$\begin{aligned}
\ddot{x} - 2\dot{y} &= \left.\frac{\partial f}{\partial x}\right|_{\tilde{x},\tilde{y}} \xi + \left.\frac{\partial f}{\partial y}\right|_{\tilde{x},\tilde{y}} \eta \\
\ddot{y} + 2\dot{x} &= \left.\frac{\partial g}{\partial x}\right|_{\tilde{x},\tilde{y}} \xi + \left.\frac{\partial g}{\partial y}\right|_{\tilde{x},\tilde{y}} \eta
\end{aligned}$$

where

$$\begin{aligned}
f(x,y) &= x - \mu(x-1+\mu)[(x-1+\mu)^2 + y^2]^{-3/2} \\
&\qquad -(1-\mu)(x+\mu)[(x+\mu)^2 + y^2]^{-3/2} \\
g(x,y) &= y - \mu y[(x-1+\mu)^2 + y^2]^{-3/2} - (1-\mu)y[(x+\mu)^2 + y^2]^{-3/2}
\end{aligned}$$

and we have used the fact that $f(\tilde{x}, \tilde{y}) = g(\tilde{x}, \tilde{y}) = 0$ to remove these terms from the equations.[1]

Using the relationships (7.9,7.10,7.11) we can re-write this as

$$\left\{ \begin{array}{c} \ddot{\xi} \\ \ddot{\eta} \end{array} \right\} = \left[\begin{array}{cc} 0 & 2 \\ -2 & 0 \end{array} \right] \left\{ \begin{array}{c} \dot{\xi} \\ \dot{\eta} \end{array} \right\} + \left[\begin{array}{cc} a & b \\ c & d \end{array} \right] \left\{ \begin{array}{c} \xi \\ \eta \end{array} \right\}$$

where

$$a = \left.\frac{\partial f}{\partial x}\right|_{\tilde{x},\tilde{y}} \qquad b = \left.\frac{\partial f}{\partial y}\right|_{\tilde{x},\tilde{y}}$$

$$c = \left.\frac{\partial g}{\partial x}\right|_{\tilde{x},\tilde{y}} \qquad d = \left.\frac{\partial g}{\partial y}\right|_{\tilde{x},\tilde{y}}$$

And this we write in the generic form

$$\ddot{\mathbf{r}} = A\dot{\mathbf{r}} + B\mathbf{r}$$

Without going into the algebra, we present the solution for L_4. The position of the equilibrium point is given by

$$\tilde{x} = \frac{1}{2} - \mu \qquad \tilde{y} = \frac{\sqrt{3}}{2} \qquad r_1 = r_2 = 1$$

[1] Note that the full equations of motions have to be used when taking the derivatives. To be completely correct, we should also retain the $\ddot{z}$ equation, and the dependence of the other equations on z. This would serve to increase the complexity without significantly altering the results, however. The dominant considerations from the viewpoint of system stability occur in the xy plane.

We take the indicated partial derivatives and plug in the values for L_4 and get

$$B = \begin{bmatrix} \frac{3}{4} & \frac{3\sqrt{3}}{2}\left(\frac{1}{2}-\mu\right) \\ \frac{3\sqrt{3}}{2}\left(\frac{1}{2}-\mu\right) & \frac{9}{4} \end{bmatrix}$$

The eigenvalue equation then becomes

$$\begin{aligned} &|\lambda^2 I - \lambda A - B| \\ &= \left| \begin{bmatrix} \lambda^2 & 0 \\ 0 & \lambda^2 \end{bmatrix} - \begin{bmatrix} \lambda \cdot 0 & 2\lambda \\ -2\lambda & \lambda \cdot 0 \end{bmatrix} - \begin{bmatrix} \frac{3}{4} & \frac{3\sqrt{3}}{2}\left(\mu-\frac{1}{2}\right) \\ \frac{3\sqrt{3}}{2}\left(\mu-\frac{1}{2}\right) & \frac{9}{4} \end{bmatrix} \right| \\ &= \begin{vmatrix} \lambda^2 - \frac{3}{4} & -2\lambda - \frac{3\sqrt{3}}{2}\left(\mu-\frac{1}{2}\right) \\ 2\lambda - \frac{3\sqrt{3}}{2}\left(\mu-\frac{1}{2}\right) & \lambda^2 - \frac{9}{4} \end{vmatrix} = 0 \end{aligned}$$

Solving for the determinant produces the quartic polynomial

$$|\lambda^2 I - \lambda A - B| = \lambda^4 + \lambda^2 - \frac{27}{4}\mu(\mu - 1) = 0$$

Treating this as a quadratic in λ^2 gives that the eigenvalues must satisfy

$$\lambda^2 = \frac{1}{2}\left[-1 \pm \sqrt{1 - 27\mu(1-\mu)}\right] \tag{7.12}$$

Now, we require for stability that all of the eigenvalues have non-positive real parts. Since this system is undamped (that is, it has no mechanism by which energy is dissipated), we cannot expect that it will be asymptotically stable. Instead, we will settle for *neutral* stability. Eigenvalues corresponding to neutral stability have zero real part, and thus when squared are negative real numbers. Since the equation we have derived is in λ^2, we look for negative real roots of eqn. (7.12).

To get such roots, it is necessary that

$$0 \le 1 - 27\mu(1-\mu) \le 1$$

since otherwise the value of λ^2 will be either imaginary or positive. Since μ and $1-\mu$ are both positive, it is clear that the $\le 1 - 27\mu(1-\mu) \le 1$ part will be satisfied. Whether the other requirement is satisfied depends upon the particular value of μ.

To find the critical values, we set

$$1 - 27\mu(1-\mu) = 0 \quad \Longrightarrow \quad \mu^2 - \mu + 1/27 = 0$$

The roots of the quadratic are approximately $\mu = 0.03852$ and $\mu = 0.96148$. The largest value of μ for any two-body pair in our solar system is that of the Earth-Moon system; the Earth is 81.3 times the mass of the moon, and $\mu = 0.0121$. Plugging in this value shows that the eigenvalues are purely imaginary, and thus the L_4 Lagrange point is neutrally stable. Since the L_5 point is an identical reflection of the L_4, it too is stable.

A similar analysis could be carried out on the three X-axis Lagrange points. It turns out that for any value of μ, these points are unstable.

In writing the equations for the three-body problem, we ignored all other bodies. In general, this is not a terrible oversight; there are few truly massive bodies in the solar system, and the distances between them are huge. Also, most of the three-body systems would consist of the Sun, a planet, and some small mass. This is not the case for the Earth-Moon system, however, and in this case leaving out the effects of the Sun is a significant error. It turns out that when the Sun is included, the Earth-Moon Lagrange points are not stable, and a body placed at one of them will not remain. This is why the Lagrange points of the Earth-Moon system are empty.

7.4 Jacobi's Integral

In this section, we will derive an analogue of the energy equation of the two-body system, which will be valid for the three-body system. We will then be able to use this constant of the motion for qualitative analysis of the motion of the infinitesimal body within the rotating frame of the three-body system.

7.4.1 A Constant of the Motion

Recall that the equations of motion of the negligible mass in the circular restricted three-body problem can be derived as (7.1-7.3)

$$\begin{aligned}
\ddot{x} - 2\dot{y} - x &= \frac{-(x-1+\mu)\mu}{r_1^3} - \frac{(x+\mu)(1-\mu)}{r_2^3} \\
\ddot{y} + 2\dot{x} - y &= \frac{-\mu y}{r_1^3} - \frac{(1-\mu)y}{r_2^3} \\
\ddot{z} &= \frac{-\mu z}{r_1^3} - \frac{(1-\mu)z}{r_2^3}
\end{aligned}$$

We have already discussed the equilibrium solutions to these equations, and noted that the linear Lagrange points are unstable, while (for appropriate values of the mass ratio μ) the triangular points are stable. These results are borne out by observation; there are asteroids at the L_4 and L_5 points in the Sun-Jupiter system. There is nothing at the triangular points of the Earth-moon system, but this is because the effects of the Sun (which are not included in the simplified model we have analyzed) render these points unstable.[2] We consider now an additional approach to analyzing the system.

[2] Recall that these points are only *neutrally* stable, so that it is relatively easy to render them unstable in actuality by imposing outside perturbations on bodies there.

In looking at the two-body problem, it was instructive to to find constants of the motion; that is, functions of the position and velocity of the body "in orbit" that remained constant at all points on the orbit. The specific energy and specific angular momentum are such constants, and provide much of the machinery for solving problems in two-body systems. It turns out that there is only one such constant for the three-body problem, discovered by Jacobi in 1836, and known as *Jacobi's Integral*.

There are several approaches to generating Jacobi's integral, and we will take one with some mechanical justification. Recalling that

$$\frac{d}{dt}\left(\frac{v^2}{2}\right) = \frac{1}{2}\frac{d}{dt}\left(\dot{\mathbf{r}}\cdot\dot{\mathbf{r}}\right) = \frac{1}{2}\left(\ddot{\mathbf{r}}\cdot\dot{\mathbf{r}}+\dot{\mathbf{r}}\cdot\ddot{\mathbf{r}}\right) = \ddot{\mathbf{r}}\cdot\dot{\mathbf{r}},$$

multiply equations (7.1) through (7.3) by $\dot{x}$, $\dot{y}$, and $\dot{z}$, respectively, and sum the results. The first terms sum to the desired dot product, and performing the algebra and simplifying, we get

$$\dot{x}\ddot{x}+\dot{y}\ddot{y}+\dot{z}\ddot{z} = x\dot{x}+y\dot{y}+\frac{\mu}{r_1^3}\left(\dot{x}-\mu\dot{x}-x\dot{x}-y\dot{y}-z\dot{z}\right)+\frac{1-\mu}{r_2^3}\left(-x\dot{x}-\mu\dot{x}-y\dot{y}-z\dot{z}\right) \tag{7.13}$$

The terms on the left comprise the derivative of $v^2/2$, and the first two terms on the right are the derivative of $(x^2+y^2)/2$. The remaining terms are not so obvious, until one recalls that

$$r_1 = \left[(x-1+\mu)^2+y^2+z^2\right]^{1/2} \qquad \text{and} \qquad r_2 = \left[(x+\mu)^2+y^2+z^2\right]^{1/2}.$$

It then becomes obvious that these terms are the time derivatives of $-\mu/r_1$ and $-(1-\mu)/r_2$, respectively. Assembling the results, and noting that there will also be a constant of integration, we have Jacobi's integral

$$C = \frac{v^2}{2}-\frac{x^2+y^2}{2}-\frac{\mu}{r_1}-\frac{1-\mu}{r_2}. \tag{7.14}$$

This is the equivalent of the energy integral for this system. As with the energy for two-body orbits, this value (sometimes called the *Jacobi constant*) is constant for any particular trajectory in the three-body problem.

Consider the terms in (7.14). The first is clearly a kinetic energy term, while the two right-most are the same type of gravitational potential terms as were derived for the two-body problem (one for each of the two massive bodies). The additional term is the square of the distance from the axis about which the entire system is rotating. This term is grouped with the velocity term. Kinetic energy depends on the velocity relative to an inertial frame; the velocity term is relative to the rotating frame. Recalling that the frame rotates at one radian per second, it becomes clear that the second term is the correction to the velocity to convert it to inertial. Therefore, the Jacobi integral is precisely an energy integral, expressed in the rotating frame.

7.4.2 Surfaces of Zero Velocity

Consider the Jacobi integral as it might apply to a vehicle in motion in the Earth-moon system. The mass of the moon is about 1/81.4 that of the earth, so

$$\mu + 81.4\mu = 1 \quad \Longrightarrow \quad \mu \approx 0.01214$$

for this system.[3] If the vehicle is very close to the surface of the earth, r_2 is very small and r_1 approaches one. Further, y must be very small (it can be no more than r_2), and x is very nearly equal to $-\mu$ (since the earth is distance μ to the left of the origin). To generate exact numbers, suppose the vehicle is at 7000 km from the center of the earth, and is exactly on the x axis between the earth and the moon. Then (recall the average radius of the orbit of the moon about the earth is 384,400 kilometers)

$$\begin{aligned} r_2 &= 7000/384400 = 0.01821 \\ r_1 &= 1 - r_2 = 0.98179 \\ x &= r_2 - \mu = 0.00607 \end{aligned}$$

The last three terms in (7.14) can then be evaluated, and we have that

$$C = v^2/2 - 54.26.$$

The question now becomes, given some value of the velocity, what is the value of C, and what does it mean?

Converting velocity into the normalized system is straightforward: the distance unit is $384,400\,\text{km}$, and the time unit is such that 2π time units equals the period of rotation of the system, which is (from tables) 27.322 days.[4] The one $DU/TU = 384400/375704\,\text{km/s} = 1.0231\,\text{km/s}$. Suppose the velocity is a bit more than the circular orbit velocity, say 8 km/sec. Then we have C = -23.7.

What does this mean? Mostly, it means that the vehicle will not be able to get very far from the earth. Just as the value of the energy in the two-body problem limits the maximum possible distance that a vehicle can travel, so too does the value of the Jacobi constant in the restricted three-body problem.

To find the limits of motion for any specific value of the Jacobi constant, we need only find the surfaces for which the velocity equals zero. These surfaces will

[3] The numbers from various sources disagree beyond the second significant digit; these are from Gurzadyan, *Theory of Interplanetary Flights*, Gordon and Breach, 1996.

[4] Recalling $\mu_\oplus = 398601\,\text{km}^3/\text{s}^2$, we could compute this from

$$TU = \sqrt{a^3/(GM)} = \sqrt{\frac{384400^3}{398601(1+1/81.4)}}.$$

Using this, the period actually comes out to 27.285 days. But, close enough.

partition space into *accessible regions*, in which v^2 is positive, and the inaccessible regions in which velocity would have to be imaginary. To characterize these regions, we will investigate the situation in which $z = 0$. However, we should keep in mind that nothing in our analysis has restricted us to the x-y plane.

Consider now the curves that are generated by setting $v = 0$ for the value of C that we have computed. Because the magnitude is very large, there are three ways that the value can be met:

1. r_2 small. Note that if r_2 is very small, then y must also be small (since it can be no larger than r_2). Also, x must be small, since its largest possible magnitude is only $\mu + r_2$. Since x and y are squared, we can ignore that term. r_1 is nearly unity, because it is the distance from the vehicle to the other primary, which is by definition unit distance from the first primary. Therefore, all of the terms other than $(1 - \mu)/r_1$ are small. In fact, we can write

   $$-C = 23.7 \approx \mu + (1 - \mu)/r_2$$

   This means that there will be a nearly circular curve of small radius about the earth, on which $v = 0$ for the given value of C. Inside this circle the velocity is positive, and thus the circle encloses an accessible region.

2. r_1 small. The analysis follows much as in the last case, except that here, $x \approx 1$ as well. Still, the term $(x^2 + y^2)/2$ will be about one-half, and since both $(1 - \mu)$ and r_1 will be nearly unity, their ratio will be about one. Thus,

   $$-C = 23.7 \approx 1/2 + 1 + \mu/r_1.$$

   Since μ is less than 1/80 times $(1 - \mu)$, it follows that there is a circle of much smaller radius about the moon, inside of which is also an accessible region.

3. Finally, it could be that the sum $x^2 + y^2$ is large. In this case, both r_1 and r_2 will also be large, and since they are divided into terms of less than one, those terms have small magnitude. Thus, there is a curve of large radius about the center of the system, outside of which is a third accessible region.

For the Earth-moon system, using the values we have chosen, the first two accessible regions are those for which $r_2 < 0.0417\,DU$ about the Earth, or for which $r_1 < 0.00055\,DU$ about the moon. The third is for distances greater than about $6\,DU$ from the center (note that none of these curves are actually circles, so these radii are only rough approximations).

The important point is that since the vehicle started out inside the first accessible region, it will stay within that region. A vehicle cannot cross a surface of zero velocity without applying thrust, so all points of any unpowered trajectory trajectory must lie within the same accessible region.

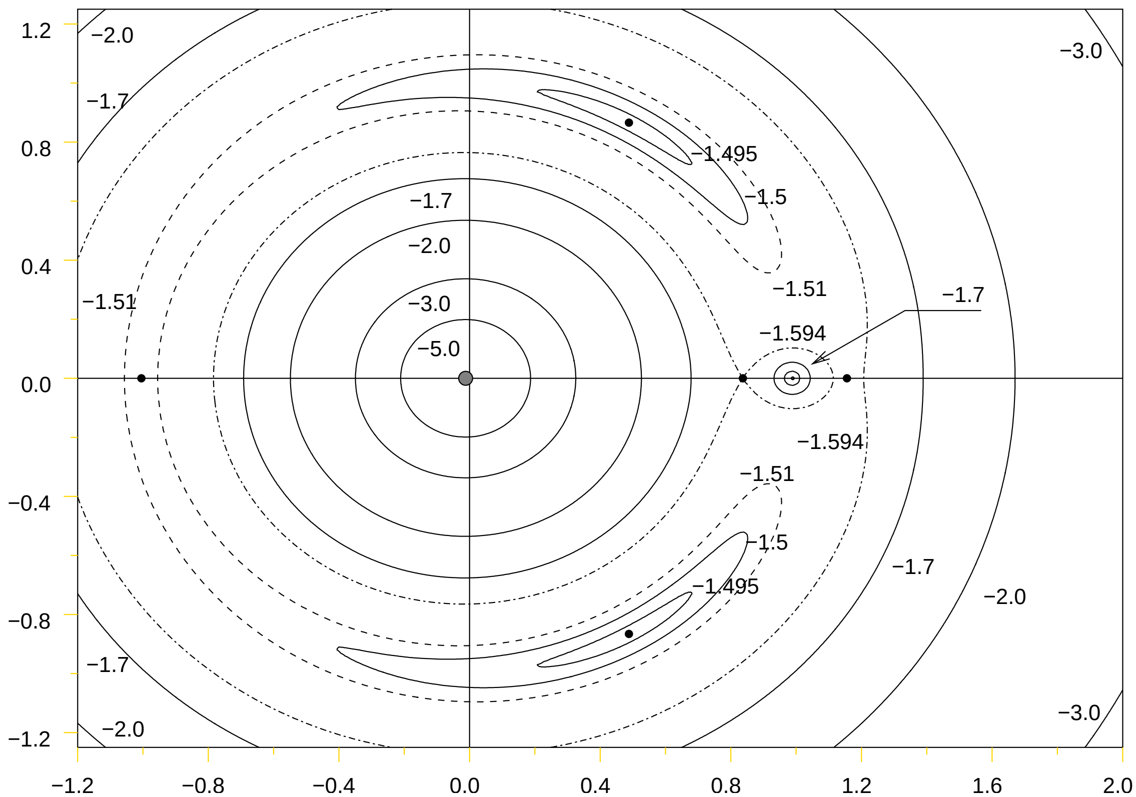


Figure 7.4: Curves of zero velocity for the Earth-moon system.

Figure 7.4 shows several curves of zero velocity for the Earth-moon system. These curves are all in the x-y plane. The lowest value represented is -5; the highest is -1.495. The earth is the shaded ball near the origin, and is slightly larger than true scale. The moon is too small to see, at about $x = 0.99$ on the x axis. The five black dots, three on the x axis and two at about $x = 0.49, y = 0.87$ are the Lagrange points. As a point of interest, the positions of the x-axis Lagrange points are about

$$x_1 = 1.156, \qquad x_2 = 0.837, \qquad x_3 = -1.005.$$

It is easy to see the way the curves move as the value of C increases towards zero. For $C = -5$ (a large magnitude), the outer curve mentioned as case 3 above is too large to fit in the sketch. The curve for case 2, about the moon, is too small to discern. The curve for case 1, about the Earth, is of substantial size. The accessible regions are inside those curves, and outside the large, off-the-chart curve.

As the value gets larger, the radius of the outer curve shrinks, so that at $C = -3$ it just cuts inside the corners of the graph on the right. The curve about the earth is larger, as is the one about the moon, though that one is still too small to see clearly. The trend continues as C increases to -2 and -1.7.

The curves for $C = -1.59411$ are shown dash-dot to set them off from the rest. Note that the curves about the earth and moon for this value meet at the L_2 point. This is the lowest value for which the two curves meet, and therefore it is the lowest value for which a craft can go from the earth to the moon; above this value, the curves merge and the earth and moon are in a single accessible region. Note that the outer curve for this value has moved in quite close, but still does not meet the inner curves; this means that a vehicle could not escape the earth-moon system.

The next lowest value shown is $C = -1.51$, and this curve is dashed. At this value, the outermost curve has blended with the inner. Note that the inaccessible region now includes both triangular points, and curves around the earth on the side away from the moon. The path to deep space is open, but it leads past the moon.

As C increases still more, this region splits into two, each of which collapses towards one of the triangular Lagrange points. The value of C at these points is -1.494. This is the highest possible value for this system; to have a higher value of the Jacobi integral, a vehicle must have a non-zero velocity.

The analysis above is done purely in the plane, but the vehicle is not restricted to the plane (recall that z is included in the formulation). When out-of-plane motion is included, the curves become *surfaces* of zero velocity; the analysis is the same, but we discuss spheres rather than circles. For example, the accessible regions for large negative values of C become the inside volumes of spheres about the earth and moon, and the volume of space outside a large sphere surrounding the entire system.

CHAPTER 8

Gravitational Presence of the Earth

To this point, we have considered all of the planets to be perfect, radially homogenous spheres. As such, we have considered them to be point masses. This is an approximation, more or less correct for different planets. In this chapter, we will explore a more accurate description of the Earth. As a result, we will also consider some corrections to the earlier results for motion about the Earth, due to terms we ignored in that earlier work.

The first section of this chapter briefly reviews the theory of conservative fields and potential functions, with emphasis on gravitational potentials. This can be skipped if desired; while it is useful for a deeper understanding of gravitational fields, it is not necessary for the implementation that follows. The second section will introduce the use of potentials for the derivation of the equations of motion. Again, this is not completely necessary for consideration of the effects of the disturbing potentials that are described in following sections.

8.1 Gravitational Potentials

The energy integral has played a large part in the analyses discussed in this book. The constant energy is the sum of two terms, and in Chapter 2 one of these terms was described as the potential energy. In this section, we will expand upon this idea. We will start by defining more rigorously the ideas of conservative fields and potentials. Then, we will discuss the gravitational potential of real, finite bodies.

8.1.1 Conservative Force Fields

The idea of a *force field* is simple enough. Given some position $\mathbf{r}$ in three-dimensional space, suppose that $\mathbf{f}(\mathbf{r})$ is the force that would be experienced by a particle with a unit amount of some characteristic quantity at that point (this can be referred to as the *specific force* in this application). If in fact the force is always known so long as the position is known, then this is a force field. The coordinate system does not have to be inertial, and in general is not.

Such fields are obviously not uncommon. Gravity is an obvious example. In this case, the characteristic quantity is mass. In an electric field, the characteristic quantity is the charge on the particle. In a magnetic field, the characteristic quantity would be related to the material of which the particle is made.

An important point to note is that the force at some point must be independent of the path leading to that point. Also, the force must depend only on the position itself, not on the velocity of the particle. That is, the force depends on where it is, not how it got there or how fast it's going.

In general, the total force on a body is the sum of many individual forces. Some of these may be due to force fields, some of them not. As an example, the forces on a ball falling through the atmosphere include weight, which is due to gravity and therefore a field force, and air resistance, which is dependent on the velocity, and is not a field force.

A force field is *conservative* if the amount of work done by the field in moving from one point to another is independent of the path the particle takes. To make this clear, consider a couple of examples.

Example 8.1. Consider the usual flat-earth approximation of motion near the surface of the planet. For motion that is small relative to the diameter of the planet, we model the surface as a plane, and model gravity as a constant, acting normal to the plane. The force on a particle of mass m is then

$$\mathbf{f}(\mathbf{r}) = -mg\hat{k} \tag{8.1}$$

Where g is the gravitational acceleration.

Now consider the work done by the field on a particle as it moves from point $\mathbf{a}$ to $\mathbf{b}$. This is a simple integral,

$$W_{\mathbf{a}\to\mathbf{b}} = \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{f}(\mathbf{r}) \cdot d\mathbf{r} = -mg \int_{\mathbf{a}}^{\mathbf{b}} dz = mg(a_k - b_k). \tag{8.2}$$

Clearly, the result is dependent only on the difference heights of $\mathbf{a}$ and $\mathbf{b}$, and is therefore independent of the path taken. ♠

Example 8.2. Consider a two-dimensional example. Let the force be defined by the function

$$\mathbf{f}(\mathbf{r}) = -ky\hat{\imath} \tag{8.3}$$

where k is some non-zero constant.

Consider now two possible paths from the origin to the point $(1, 1)$. In the first, we travel along the x axis from $(0, 0)$ to $(1, 0)$ and then in the $\hat{\jmath}$ direction from $(1, 0)$ to

$(1,1)$. The work done by the field is

$$W_1 = -k\int_{(0,0)}^{(1,0)} 0\hat{\imath}\cdot\hat{\imath} - k\int_{(1,0)}^{(1,1)} y\hat{\imath}\cdot\hat{\jmath} = 0. \tag{8.4}$$

Note that the first term is zero because $y = 0$ along that segment, so that the magnitude of the force is zero.

The second path will run from the origin along the y axis to $(0,1)$, and then in the $\hat{\imath}$ direction to $(1,1)$. The work done by the field along this path is

$$W_2 = -k\int_{(0,0)}^{(1,0)} y\hat{\imath}\cdot\hat{\jmath} - k\int_{(1,0)}^{(1,1)} (1)\hat{\imath}\cdot\hat{\imath} = -k. \tag{8.5}$$

The two paths yield different values of work, so the field is not conservative. ♠

There are a few ways to mathematically express characteristics of a field that guarantee that it is conservative. One is that the work along any closed path is zero; this is a simple consequence of path independence, and difficult to prove directly. But a very simple way is to use the integral of the work along any path from some arbitrary point $\mathbf{r}_1$ to another, $\mathbf{r}_2$. Let the force be defined as

$$\mathbf{f}(\mathbf{r}) = f_i\hat{\imath} + f_j\hat{\jmath} + f_k\hat{k}. \tag{8.6}$$

Letting $d\mathbf{r}$ be written as $dx\,\hat{\imath} + dy\,\hat{\jmath} + dz\,\hat{k}$, we have

$$\mathbf{f}(\mathbf{r})\cdot d\mathbf{r} = f_i dx + f_j dy + f_k dz. \tag{8.7}$$

Now suppose that we can find a scalar function $\Phi(\mathbf{r})$ such that

$$f_i = \frac{\partial\Phi}{\partial x}, \qquad f_j = \frac{\partial\Phi}{\partial y}, \qquad f_k = \frac{\partial\Phi}{\partial z}. \tag{8.8}$$

Then the work integral becomes

$$W = \int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{f}(\mathbf{r})\cdot d\mathbf{r} = \int_{\mathbf{r}_1}^{\mathbf{r}_2}\left(\frac{\partial\Phi}{\partial x}dx + \frac{\partial\Phi}{\partial y}dy + \frac{\partial\Phi}{\partial z}dz\right) = \Phi(\mathbf{r}_2) - \Phi(\mathbf{r}_1). \tag{8.9}$$

We see therefore that if such a function exists, then the field *must* be conservative, since the value of the integral will always be defined by the values of the function at the endpoints, regardless of the path traversed.

Recalling section 2.3.1, we see that this is very similar to the potential energy defined there. In fact, we have

$$\Phi(\mathbf{r}_2) - \Phi(\mathbf{r}_1) = \int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{f}(\mathbf{r})\cdot d\mathbf{r} = \int_{\mathbf{r}_1}^{\mathbf{R}} \mathbf{f}(\mathbf{r})\cdot d\mathbf{r} - \int_{\mathbf{r}_2}^{\mathbf{R}} \mathbf{f}(\mathbf{r})\cdot d\mathbf{r} = U(\mathbf{r}_1) - U(\mathbf{r}_2). \tag{8.10}$$

This implies that the change in Φ over any path is the negative of the change in the potential energy.

Now, we note that adding a constant value to Φ will not change its derivatives. We can therefore choose $\Phi(\mathbf{r})$ to be the negative of the potential energy function $U(\mathbf{r})$. In summary, we can state from this analysis that a conservative force field is one for which a potential energy function $U(\mathbf{r})$ can be defined, and further, the specific force on a particle due to the field is given by the negative gradient of the potential energy, as

$$f_i = -\frac{\partial U}{\partial x}, \qquad f_j = -\frac{\partial U}{\partial y}, \qquad f_k = -\frac{\partial U}{\partial z}. \tag{8.11}$$

The function $\Phi(\mathbf{r}) = -U(\mathbf{r})$ is often known as the *potential function.* In many cases, the distinction between the two is ignored, and U is referred to as the potential function. We further make the declaration that *the existence of a potential function (equivalently, a potential energy function) is sufficient to prove that a field is conservative.*

It has been easy to show that the existence of a potential function is enough to prove that the field is conservative. It is also true that if such a function does not exist, then the field is not conservative.

Example 8.3. Recall the field of example 8.2,

$$\mathbf{f}(x, y) = -ky\hat{\imath}.$$

Show that it is not conservative.

Recall that for *any* function $\Phi(x, y)$ that is twice differentiable, it must be true that

$$\frac{\partial^2 \Phi}{\partial x \partial y} = \frac{\partial^2 \Phi}{\partial y \partial x}.$$

Therefore, if there is a potential function for this field, then it must be (from the definition) true that

$$f_i = \frac{\partial \Phi}{\partial x} \quad \text{and} \quad f_j = \frac{\partial \Phi}{\partial y},$$

and therefore it follows that

$$\frac{\partial f_i}{\partial y} = \frac{\partial f_j}{\partial x}.$$

In this case, this requires

$$\frac{\partial(-ky)}{\partial y} = \frac{\partial 0}{\partial x} \quad \Longrightarrow \quad -k = 0$$

which is clearly nonsense. Therefore no potential function exists, and the field is thus not conservative. ♠

Since the force is the gradient of the potential function, the field is fully defined by the potential function.

8.1.2 Gravitational fields

To begin, realize that Newton's law of gravitation applies only to particles. There is no consideration of the size of the body, or of the distribution of mass within the body. To extend the law of gravitation to a body of finite size and possibly non-uniform mass distribution, it is necessary to integrate over the body. In this section, we will assume that the bodies have uniform mass distribution.

First, however, consider the potential of a particle. We have from earlier that

$$\Phi(\mathbf{r}) = \frac{\mu}{r} \tag{8.12}$$

Letting the elements of $\mathbf{r}$ be x, y, and z, and noting that

$$r = \sqrt{x^2 + y^2 + z^2},$$

we have

$$\mathbf{a} = \nabla\Phi = \left\{ \begin{array}{c} \partial\Phi/\partial x \\ \partial\Phi/\partial y \\ \partial\Phi/\partial z \end{array} \right\} = \frac{1}{r^3} \left\{ \begin{array}{c} -\mu x \\ -\mu y \\ -\mu z \end{array} \right\} = -\frac{\mu \mathbf{r}}{r^3} \tag{8.13}$$

just as we require. Here, we have replaced the specific force, in this case the force per unit mass, with the acceleration.

Now suppose we have two massive particles providing gravitational potentials. The total potential is just the sum of the two:

$$\mathbf{a} = \frac{\partial Phi}{\partial \mathbf{r}} = \frac{\partial}{\mathbf{r}} \left(\Phi_1 + \Phi_2\right) = \frac{\partial \Phi_1}{\partial \mathbf{r}} + \frac{\partial \Phi_2}{\partial \mathbf{r}}.$$

It follows that we can integrate over an infinite number of potentials, which provides the justification for what follows in this section.

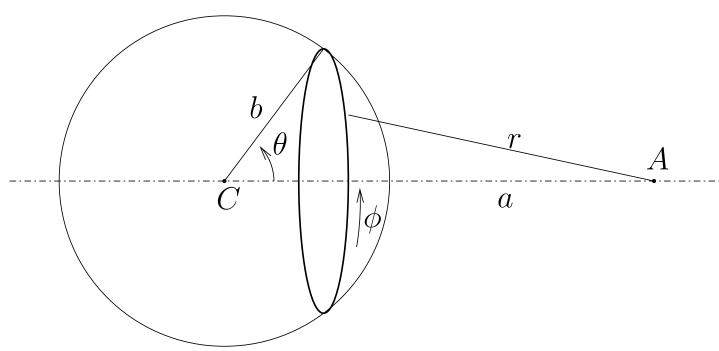


Figure 8.1: Integration over annulus.

As an example, consider the gravitational potential that would be created by a uniform hollow spherical shell, at a point A outside of the shell. Let the shell have uniform density ρ, radius b, thickness db and be centered at the point C. At any point on the annulus sketched in Figure 8.1, the differential mass is

$$dm = \rho(bd\theta)(b\sin\theta d\phi)db.$$

The distance from the point on the annulus to A is

$$r^2 = a^2 + b^2 - 2ab\cos\theta,$$

where a is the distance from C to A. The potential of the annulus at A is then the integral over ϕ of the potential of the differential mass:

$$\Phi_A = \frac{G\rho b^2 \sin\theta d\theta db}{(a^2 + b^2 - 2ab\cos\theta)^{1/2}} \int_0^{2\pi} d\phi = \frac{2\pi G\rho b^2 \sin\theta d\theta db}{r}.$$

The potential of the shell is the integral of the potential of the annuluis over θ from 0 to π. Noting that the mass of the shell is

$$M_s = 4\pi\rho b^2 db,$$

this is

$$\Phi_S = 2\pi G\rho b^2 db \int_0^{\pi} \frac{\sin\theta d\theta}{(a^2 + b^2 - 2ab\cos\theta)^{1/2}} = \frac{GM_S}{a}. \tag{8.14}$$

This is precisely the potential of a particle of mass M_s located at C.

It is clear that to find the potential of a uniform sphere, we would integrate the result in (8.14) over the radius, and the result would be the same. Thus, from any point outside of a sphere, the gravitational potential is the same as that of a point containing all of the mass of the sphere, located at the center of the sphere. This result holds as well for a sphere in which the density varies as a function of distance from the center, so long as it is homogenous within any shell.

Note that this is a formal justification of the approximation made in chapter 2, where we treated all bodies as points. To the extent that the Earth and other planets can be treated as radially symmetric spheres, the approximation is valid, even for points very close to the surface. This is a very good approximation for most purposes, though the errors introduced are significant in many applications.

8.2 Equations of Motion using Potentials

In this section, the equation of motion of a mass in the presence of more than one additional mass will he derived, from the point of view of gravitational potentials. This will be similar to the sphere of influence (section 6.2), from a different and more general approach.

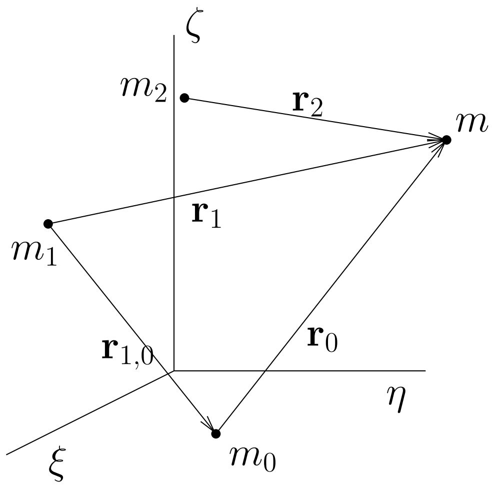


Figure 8.2: The n-body problem.

Figure 8.2 shows a body m in space with three others.[1] Following the results of the last section, the potential at m due to the other bodies is

$$\Phi(\cdot) = \frac{Gm_0}{r_0} + \frac{Gm_1}{r_1} + \frac{Gm_2}{r_2} \tag{8.15}$$

Since the force per unit mass on m is the gradient of the potential, and the acceleration is (from the second law of motion) also the force per unit mass, we have

$$\frac{d\mathbf{v}}{dt} = \nabla\Phi. \tag{8.16}$$

For simplicity, consider the motion in the ξ direction. Recalling that

$$r_0^2 = (\xi - \xi_0)^2 + (\eta - \eta_0)^2 + (\zeta - \zeta_0)^2$$

we have

$$\frac{\partial}{\partial\xi}\left(\frac{1}{r_0}\right) = -\frac{\xi - \xi_0}{r_0^3}$$

with similar results for r_1 and r_2. Taking the partial derivative of the potential function then gives

$$\ddot{\xi} = \frac{\partial\Phi}{\partial\xi} = \frac{-Gm_0(\xi - \xi_0)}{r_0^3} - \frac{Gm_1(\xi - \xi_1)}{r_1^3} - \frac{Gm_2(\xi - \xi_2)}{r_2^3} \tag{8.17}$$

The expressions for $\ddot{\eta}$ and $\ddot{\zeta}$ are very similar.

Consider the motion of mass m_0. The potential at m_0 does not include a contribution due to m_0 itself, but does include a contribution due to m. Thus

$$\Phi_0(\cdot) = \frac{Gm}{r_0} + \frac{Gm_1}{r_{1,0}} + \frac{Gm_2}{r_{2,0}} \tag{8.18}$$

where we use $\mathbf{r}_{i,j}$ to mean the vector from m_i to m_j. Clearly,

$$\ddot{\xi}_0 = \frac{\partial\Phi_0}{\partial\xi_0} = \frac{-Gm_0(\xi_0 - \xi)}{r_0^3} - \frac{Gm_1(\xi_0 - \xi_1)}{r_{1,0}^3} - \frac{Gm_2(\xi_0 - \xi_2)}{r_{2,0}^3}. \tag{8.19}$$

To examine the motion of m relative to m_0, define a new coordinate system centered at m_0, with axes parallel to those of the intertial frame. Define

$$x = \xi - \xi_0, \qquad y = \eta - \eta_0, \qquad z = \zeta - \zeta_0.$$

[1]Three is a sufficient approximation to "many" in this case; the results we will derive will easily extend.

Then we have $\ddot{x} = \ddot{\xi} - \ddot{\xi}_0$, and

$$\begin{aligned}\ddot{x} \quad = \quad & \frac{-Gm_0(\xi - \xi_0)}{r_0^3} - \frac{Gm_1(\xi - \xi_1)}{r_1^3} - \frac{Gm_2(\xi - \xi_2)}{r_2^3} \\ & + \frac{Gm(\xi_0 - \xi)}{r_0^3} + \frac{Gm_1(\xi_0 - \xi_1)}{r_{1,0}^3} + \frac{Gm_2(\xi_0 - \xi_2)}{r_{2,0}^3} \end{aligned} \tag{8.20}$$

Re-arranging the terms provides

$$\ddot{x} = \frac{-G(m_0 + m)(\xi - \xi_0)}{r_0^3} - Gm_1 \left[\frac{\xi - \xi_1}{r_1^3} + \frac{\xi_1 - \xi_0}{r_{1,0}^3} \right] - Gm_2 \left[\frac{\xi - \xi_2}{r_2^3} + \frac{\xi_2 - \xi_0}{r_{2,0}^3} \right] \tag{8.21}$$

We will henceforth make the usual identification $\mu = G(m_0 + m)$.

Now, we wish to replace the variables ξ, ξ_0, et cetera with variables defined in the new coordinate frame. To do so, define

$$\begin{aligned} x_1 = \xi_1 - \xi_0, \quad y_1 = \eta_1 - \eta_0, \quad z_1 = \zeta_1 - \zeta_0, \\ x_2 = \xi_2 - \xi_0, \quad y_2 = \eta_2 - \eta_0, \quad z_2 = \zeta_2 - \zeta_0. \end{aligned}$$

$$\ddot{x} = -\frac{\mu x}{r^3} - Gm_1 \left[\frac{x - x_1}{r_1^3} + \frac{x_1}{r_{1,0}^3} \right] - Gm_2 \left[\frac{x - x_2}{r_2^3} + \frac{x_2}{r_{2,0}^3} \right] \tag{8.22}$$

Noting that

$$r_1 = \sqrt{(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2} \quad \text{and} \quad r_{1,0} = \sqrt{x_1^2 + y_1^2 + z_1^2},$$

we have that

$$\frac{x - x_1}{r_1^3} = -\frac{\partial}{\partial x} \left(\frac{1}{r_1} \right) \tag{8.23}$$

and

$$\frac{x_1}{r_{1,0}^3} = \frac{\partial}{\partial x} \left(\frac{x x_1}{r_{1,0}^3} \right). \tag{8.24}$$

Combining these with similar results for terms involving x_2 leads to

$$\ddot{x} = -\frac{\mu x}{r^3} + \frac{\partial}{\partial x} \left(\frac{1}{r_1} - \frac{x x_1}{r_{1,0}^3} \right) + \frac{\partial}{\partial x} \left(\frac{1}{r_2} - \frac{x x_2}{r_{2,0}^3} \right) \tag{8.25}$$

The corresponding relations for y and z are

$$\ddot{y} = -\frac{\mu y}{r^3} + \frac{\partial}{\partial y} \left(\frac{1}{r_1} - \frac{y y_1}{r_{1,0}^3} \right) + \frac{\partial}{\partial y} \left(\frac{1}{r_2} - \frac{y y_2}{r_{2,0}^3} \right) \tag{8.26}$$

and

$$\ddot{z} = -\frac{\mu z}{r^3} + \frac{\partial}{\partial z} \left(\frac{1}{r_1} - \frac{z z_1}{r_{1,0}^3} \right) + \frac{\partial}{\partial z} \left(\frac{1}{r_2} - \frac{z z_2}{r_{2,0}^3} \right). \tag{8.27}$$

Combining equations (8.25), (8.26), and (8.26) provides

$$\ddot{\mathbf{r}} = -\frac{\mu\ddot{\mathbf{r}}}{r^3} + \frac{\partial R_1}{\partial \mathbf{r}} + \frac{\partial R_2}{\partial \mathbf{r}} \tag{8.28}$$

where

$$R_1 = \frac{1}{r_1} - \frac{xx_1 + yy_1 + zz_1}{r_{1,0}}; \qquad R_2 = \frac{1}{r_2} - \frac{xx_2 + yy_2 + zz_2}{r_{2,0}}.$$

It should be clear that if there are any further planets, each adds a term to (8.28) of the form

$$\frac{\partial R_k}{\partial \mathbf{r}} = \frac{\partial}{\partial \mathbf{r}}\left(\frac{1}{r_1} - \frac{xx_1 + yy_1 + zz_1}{r_{1,0}}\right). \tag{8.29}$$

The functions R_k in the equations above are known as *disturbing potentials*. Without them, the equation of motion is simply

$$\ddot{\mathbf{r}} = \frac{\partial}{\partial \mathbf{r}}\left(\frac{\mu}{r}\right), \tag{8.30}$$

and the term on the right hand side is recognized as the potential due to a spherical planet. In general, the equation of motion when the system does not meet the perfect two-body assumptions can be written as

$$\ddot{\mathbf{r}} = -\frac{\mu\ddot{\mathbf{r}}}{r^3} + \frac{\partial R}{\partial \mathbf{r}} \tag{8.31}$$

where R is the sum of all disturbing potentials.

In the early chapters of this text, planets were considered to be in two-body motion about the Sun. This produced an equation of motion such as (8.30). A complete equation of motion would be of the form (8.31), where the other planets make themselves known through R.

8.3 Gravitational Potential of the Earth

At this point, we have shown that any force field caused by gravity can be expressed as the gradient of a potential function. In the last section, the potential function due to the presence of several bodies was derived, holding the assumption that the bodies can be modelled as discrete points. This resulted in (8.31), with a disturbing potential due to the existence of additional planets. Here, we consider a disturbing potential arising from a different source.

In deriving the equations of motion about a planet, we have made the assumption that the planet was spherical and radially symmetric. This is not perfectly true for any planet, though it is close.[2] The planet about which we know most, of course, and about which we are primarily concerned, is the earth. In this section, we examine the gravitational field of the earth.

We have said that the approximation of the earth as a sphere is a good one, and this is true. A better approximation is that it is an *oblate spheroid*, which is to say that it is a solid body with an axis of radial symmetry; any plane normal to the axis that cuts through the body results in a circular section, and any cutting plane that includes the axis results in an elliptic cross section. A further refinement has the spheroid slightly larger below the equator than above. And the refinements continue.

The actual shape and mass distribution of the planet is important for many purposes, but from the point of view of spaceflight only its appearance as a source of gravitational attraction is significant.[3] The gravitational field is generally described by its associated potential function. In this section, we discuss the usual forms of that this description takes.

Note: In many commercial satellite simulation computer programs, the gravitational field is actually expressed as a table based on latitude and longitude. This is specific to the purpose, however, and not germane to the current discussion. ♢

8.3.1 Legendre Polynomials

The usual expression of the potential function of the Earth involves *Legendre Polynomials.* These polynomials are one of several sets of functions

8.3.2 Expression of the Gravitational Field

As mentioned, the gravitational field of the earth is expressed through the potential function. Given enough terms, the potential of any body, of whatever shape, can be expressed to an arbitrarily high level of precision using Legendre polynomials.

[2]In 1995, a body larger than Pluto was found to be orbiting the Sun at a distance of approximately $35AU$ (the body was eventually given the name Xena). In the years since, several more were found. In response, the International Astronomical Society adopted a formal definition of a planet. Part of this definition is that the body must have sufficient mass that its own gravity forces it to be spherical. Thus, all planets will be close to perfectly spherical.

[3]With the obvious caveat that the body in flight have sufficient altitude to avoid the surface.

Chapter 9

Satellite Clusters and Relative Motion

The idea of relative motion in orbit has been of interest ever since the space age truly began in the 1950s. From the beginning, there was interest in creating space stations, and in missions to the Moon. In each of these cases, there would be a need for small craft to maneuver to, and dock with, larger craft (for lunar missions, a command module was placed in orbit around the moon, and the lander had to rendezvous with it after lifting off from the lunar surface to begin the return trip).

More recently, in the last decade of the 20^{th} century, interest has grown in *satellite clusters*. Ever since the first satellites were put in earth orbit, they have been being made larger and more complex. At the same time, the communication and navigation systems for the satellites have been made more and more capable. At some point, it became obvious that in some cases, it would make more sense to divide the duties among several smaller, less complex, less expensive satellites, which would remain near each other as they orbited the earth. There are also missions which are possible only with such clusters, and which require very precise knowledge of the relative positions of the satellites within the cluster.

We will not explore the nature of these missions, which largely involve highly accurate imaging of the earth, or of distant stars. Instead, taking it as given that clusters are desireable, we will concentrate on the nature of the relative motion of satellite in clusters. To begin, however, we will consider the most basic use of relative motion, the docking of a craft with an orbiting space station.

9.1 Rendezvous in Circular Orbit

In this section, we explore relative motion by considering the motion of a powered body attempting rendezvous with a second, unpowered body. The next section will describe the overall approach, which will then be specialized to the case in which the unpowered body is in circular orbit. We will think of these as a spacecraft and an orbiting space station.

All of the following analysis will assume that the planet is spherical. The additional accelerations due to planetary oblateness, or the presence of other bodies (such as the moon) would be nearly identical for both the spacecraft and the space station, and so very nearly cancel out. While some work has been done that includes some of these effects, they are not needed for an introduction to the problem.

9.1.1 Equations of Relative Motion

In general, the equation of motion of a spacecraft in orbit is

$$\frac{{}^I d^2}{dt^2}(\mathbf{r}) = \frac{-\mu\mathbf{r}}{r^3} + \mathbf{f} \tag{9.1}$$

where $\mathbf{f}$ is the acceleration due to thrust.[1] This equation assumes that the planet is spherical, and that there are no other forces or accelerations present. The prefixed I in (9.1) notes that the derivative is taken with respect to a fixed, or at least non-rotating, reference frame.[2]

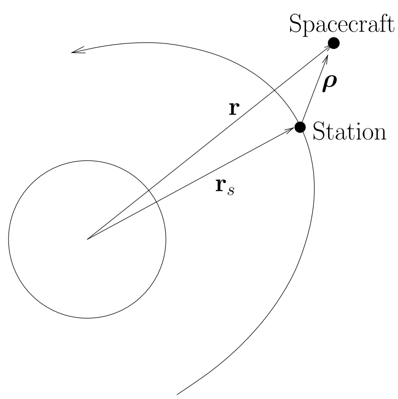


Figure 9.1: Spacecraft in flight near space station.

When the problem being considered is how the spacecraft moves when near the space station, it makes sense to think of the position of the craft relative to the station (or the astronaut relative to the shuttle, or whatever). To that end, define the position of the craft as

$$\mathbf{r} = \mathbf{r}_s + \boldsymbol{\rho} \tag{9.2}$$

where $\mathbf{r}_s$ is the position vector of the space station, $\mathbf{r}$ is the position of the spacecraft, and $\boldsymbol{\rho}$ is the relative position vector. The relative motion equation is then

$$\frac{{}^I d^2}{dt^2}(\boldsymbol{\rho}) = \frac{{}^I d^2}{dt^2}(\mathbf{r}) - \frac{{}^I d^2}{dt^2}(\mathbf{r}_s). \tag{9.3}$$

Now, using (9.1), and assuming that the station is not thrusting, so that its orbit is Keplerian, we have

$$\frac{{}^I d^2}{dt^2}(\boldsymbol{\rho}) = \frac{-\mu(\boldsymbol{\rho} - \mathbf{r}_s)}{\|\boldsymbol{\rho} - \mathbf{r}_s\|^3} + \mathbf{f} - \frac{-\mu\mathbf{r}_s}{r_s^3}. \tag{9.4}$$

[1] $\mathbf{f}$ should not be taken to be the *force* from the spacecraft motors. It might be better to use $\mathbf{a}$ which more obviously means acceleration, but the letter a appears too often as semimajor axis to give it another meaning. Instead, the reader should simply take $\mathbf{f}$ as the *specific* force, or force per unit mass, which appears as acceleration in application.

[2] Recall that the equation of motion for the two-body problem has the same form in both the inertial frame and in a non-rotating frame attached to one of the bodies.

Now, because the size of the vector $\boldsymbol{\rho}$ is much less than that of $\mathbf{r}_s$, we expand the motion of the spacecraft about that of the station. There are several ways to approach the expansion. We will use Taylor expansion, which is covered briefly in Appendix C. Renaming the terms in (9.4) and writing it as

$$\frac{{}^I d^2}{dt^2}(\boldsymbol{\rho}) = \mathbf{F}(\mathbf{r}) + \mathbf{f} - \mathbf{F}(\mathbf{r}_s), \tag{9.5}$$

we expand the expression for unthrusted motion of the spacecraft as

$$\mathbf{F}(\mathbf{r}) = \mathbf{F}(\mathbf{r}_s) + \frac{d}{d\mathbf{r}}\mathbf{F}(\mathbf{r}_s)\,\boldsymbol{\rho} + \cdots \tag{9.6}$$

The terms beyond those listed involve higher powers of $\boldsymbol{\rho}$, and so will be considered negligible.

Since the first term after the equal sign is identical to the final term in (9.5), the two cancel. The result is a *first-order* approximation to the equations of relative motion

$$\frac{{}^I d^2}{dt^2}(\boldsymbol{\rho}) = \frac{d}{d\mathbf{r}}\mathbf{F}(\mathbf{r}_s)\,\boldsymbol{\rho} + \mathbf{f}. \tag{9.7}$$

This is also known as a *linear approximation* to the equations of motion, and it is said that the system has been linearized.

To write out the equations, we let the elements of $\boldsymbol{\rho}$ be $\{\xi, \eta, \zeta\}$ so that

$$\mathbf{r} = \left\{\begin{array}{c} x \\ y \\ z \end{array}\right\} = \left\{\begin{array}{c} x_s \\ y_s \\ z_s \end{array}\right\} + \left\{\begin{array}{c} \xi \\ \eta \\ \zeta \end{array}\right\}. \tag{9.8}$$

Recalling that $\mathbf{F}(\mathbf{r})$ can be written as

$$\mathbf{F}(\mathbf{r}) = -\frac{\mu}{(x^2+y^2+z^2)^{3/2}}\left\{\begin{array}{c} x \\ y \\ z \end{array}\right\}$$

the partial derivatives of the first of these becomes

$$\begin{aligned}
\frac{\partial F_1}{\partial x} &= -\frac{\mu}{(x^2+y^2+z^2)^{3/2}} + \frac{3\mu x^2}{(x^2+y^2+z^2)^{5/2}} = -\frac{\mu}{r^3} + \frac{3\mu x^2}{r^5} \\
\frac{\partial F_1}{\partial y} &= \frac{3\mu xy}{(x^2+y^2+z^2)^{5/2}} = \frac{3\mu xy}{r^5} \\
\frac{\partial F_1}{\partial z} &= \frac{3\mu xz}{(x^2+y^2+z^2)^{5/2}} = \frac{3\mu xz}{r^5}.
\end{aligned}$$

Continuing the derivation produces the system

$$\frac{d}{d\mathbf{r}}\mathbf{F}(\mathbf{r}_s) = \frac{\mu}{r^5}\left(3\mathbf{r}\mathbf{r}^T - I_3 r^2\right) \tag{9.9}$$

(Here, I_3 means the 3×3 identity matrix). At this point, the only simplifications have been to adopt the two-body assumptions.

9.1.2 Specialization to Circular Orbits

In this section we assume that the space station is in a circular orbit about the earth. We also define a reference frame in which to define the vectors $\mathbf{r}_s$ and $\boldsymbol{\rho}$. As shown in Figure 9.2, we let the frame be centered at the space station, and take the $\hat{\imath}$ direction to be radially out from the earth. The $\hat{k}$ unit vector is parallel to the angular momentum vector of the station's orbit, and $\hat{\jmath}$ is chosen to complete the right-handed coordinate system.

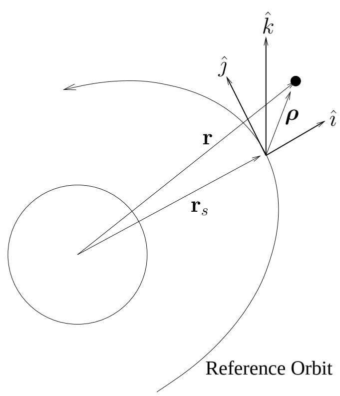


Figure 9.2: Spacecraft in flight near space station.

The defined coordinate system rotates at the angular rate at which the station is moving. Because the frame is rotating, we have the acceleration in the moving frame to be

$$\ddot{\boldsymbol{\rho}} = \frac{{}^I d^2}{dt^2}(\boldsymbol{\rho}) - 2\boldsymbol{\omega}\times\dot{\boldsymbol{\rho}} - \dot{\boldsymbol{\omega}}\times\boldsymbol{\rho} - \boldsymbol{\omega}\times\boldsymbol{\omega}\times\boldsymbol{\rho} \tag{9.10}$$

(the usual subtraction of the acceleration of the origin of the local frame is not necessary, as this was done in deriving eqn. 9.5).

Invoking the assumption of a circular orbit for the space station, the angular velocity vector becomes

$$\boldsymbol{\omega} = \left\{ \begin{array}{c} 0 \\ 0 \\ n \end{array} \right\} \tag{9.11}$$

where n is the mean motion. This is constant, so the term involving its time derivative in (9.10) disappears. Further, we have

$$\mathbf{r}_s = \left\{ \begin{array}{c} r_s \\ 0 \\ 0 \end{array} \right\}. \tag{9.12}$$

Evaluating (9.9) at $\mathbf{r} = \mathbf{r}_s$, and using the result along with (9.11) and (9.7) in (9.10), provides the result

$$\ddot{\xi} = 3n^2\xi + 2n\dot{\eta} + f_i \tag{9.13}$$

$$\ddot{\eta} = -2n\dot{\xi} + f_j \tag{9.14}$$

$$\ddot{\zeta} = -n^2\zeta + f_z. \tag{9.15}$$

where $n = \sqrt{\mu/r_s^3}$ has been used to simplify some of the terms. These are known variously as *Hill's* equations, the *Clohessy-Wiltshire* equations, and occasionally as the *Clohessy-Wiltshire-Hill* equations. They are very similar to equations derived by Hill in 1878, in his studies of the motion of the moon.[3] Clohessy and Wiltshire re-derived them in the context of orbital rendezvous in 1960.[4] We will refer to them as the CW equations, due to the context in which we are using them. These equations describe the motion of a body relative to another, where the second is in a circular reference orbit. Since the presence of the space station does not actually enter the dynamics, it need not actually exist. All that is required is that the position of the origin of hte local frame be defined as a point moving along a Keplerian reference orbit. The term *reference orbit* will often be used to describe the motion of the local origin.

9.1.3 Natural Motion

The CW equations cannot be solved in general unless the thrust $\mathbf{f}(t)$ is known. However, if impulsive maneuvers are assumed, the thrust is zero during motion. With this assumption, the equations can be solved in closed form. Motion without thrust is also known as the *natural motion* of the satellite.

The easiest equation to solve is (9.15), as neither ξ nor η appear. It is thus *decoupled* from the other two equations. With $\mathbf{f} = 0$, we have

$$\ddot{\zeta} = -n^2\zeta \quad \Longrightarrow \quad \zeta(t) = c_1 \sin nt + c_2 \cos nt \tag{9.16}$$

where the constants c_1 and c_2 depend on the initial conditions. Solving in terms of these, we have

$$\zeta(t) = \left(\dot{\zeta}_0/n\right) \sin nt + \zeta_0 \cos nt. \tag{9.17}$$

The other equations are only slightly more difficult. Integrating (9.14) gives

$$\dot{\eta} = -2n\zeta + k \tag{9.18}$$

where k is a constant of integration. This is substituted into (9.13) to get

$$\ddot{\xi} = -n^2\xi + 2nk \tag{9.19}$$

which is solved directly to get

$$\xi(t) = c_3 \sin nt + c_4 \cos nt + 2k/n. \tag{9.20}$$

We substitute (9.20) into (9.18) and integrate to get

$$\eta(t) = 2c_3 \cos nt - 2c_4 \sin nt - 3kt + c_5. \tag{9.21}$$

[3] G. W. Hill, "Researches in the Lunar Theory", *American Journal of mathematics*, vol. 1, 1878.

[4] W. H. Clohessy and R. S. Wiltshire, "Terminal Guidance Systems for Satellite Rendezvous", *Journal of the Aerospace Sciences*, Sept. 1960

Solving for the constants in terms of the initial conditions produces finally

$$\xi(t) = \frac{\dot{\xi}_0}{n}\sin nt - \left(\frac{2\dot{\eta}_0}{n} + 3\xi_0\right)\cos nt + \frac{2}{n}\left(\dot{\eta}_0 + 2n\xi_0\right) \tag{9.22}$$

$$\eta(t) = \frac{2\dot{\xi}_0}{n}\cos nt + \left(\frac{4\dot{\eta}_0}{n} + 6\xi_0\right)\sin nt - 3\left(\dot{\eta}_0 + 2n\xi_0\right)t - \frac{2\dot{\xi}_0}{n} + \eta_0 \tag{9.23}$$

It is interesting to consider certain special cases of the motion. From (9.23), for instance, it can be seen that if all initial conditions other than η_0 are zero, then $\xi(t) \equiv 0$ and $\eta(t) \equiv \eta_0$. This means that it is possible to simply offset the spacecraft by some distance ahead of or behind the space station, and it will remain in that relative position. It is not possible, however, to do the same with a non-zero value of ξ_0. Any offset in the radial direction will result in motion relative to the station.

Another point of interest is the term multiplying time in (9.23). A term in the solution of a physical system that grows with time is called a *secular* term. In this case, the coefficient is $-3(\dot{\eta}_0 + 2n\xi_0)$. If this term is non-zero, then the spacecraft will drift away from the station in the along-track direction, pulling ahead or falling behind depending on the sign of the coefficient. Note also that if this coefficient is zero, the constant term in (9.22) is also zero.

Considering (9.22) more closely, it is seen to be a simple oscillation, with a constant offset from zero. If the secular term is zero, then this offset is also zero.

How does this relate to what we already know about orbits? If the spacecraft continually pulls ahead of the station, then in the time the station makes one orbit, the craft will already have completed an orbit. This means that the orbital period of the craft is different, which in turn means that the orbital energy and the semimajor axis are different. It is this difference in the semimajor axis that causes the offset in the radial direction. We refer to the requirement that

$$\dot{\eta}_0 + 2n\xi_0 = 0 \tag{9.24}$$

for no secular growth to occur as the *energy matching* condition.

If the energy matching condition is satisfied, the equations simplify. Using (9.24), the coefficients of $\cos nt$ in (9.22) and of $\sin nt$ in (9.23) become ξ_0 and $-2\xi_0$, respectively. Equations (9.22) and (9.23) then simplify to

$$\xi(t) = (\dot{\xi}_0/n)\sin nt + \xi_0\cos nt \tag{9.25}$$

$$\eta(t) = 2(\dot{\xi}_0/n)\cos nt - 2\xi_0\sin nt - 2(\dot{\xi}_0/n) + \eta_0. \tag{9.26}$$

The constant η_0 has no effect on the form of the motion, and can be safely ignored for now.

Recall that the form of the equation in (9.25) can be re-written as

$$k_1\sin nt - k_2\cos nt = A_\xi\cos(nt + \psi).$$

where a_ξ is the amplitude and ψ a constant phase angle. Also, recalling that

$$\cos\alpha = -\sin(\alpha - \pi/2); \qquad \sin\alpha = \cos(\alpha - \pi/2)$$

the equations become

$$\begin{aligned} \xi(t) &= A_\xi \cos(nt + \psi) && (9.27)\\ \eta(t) &= -2A_\xi \cos(nt + \psi - \pi/2) - 2(\dot{\xi}_0/n) + \eta_0. && (9.28) \end{aligned}$$

It becomes clear that the relations for $\xi(t)$ and $\eta(t)$ are sinusoids that are ninety degrees out of phase, with the amplitude of η twice that of ξ. In other words, the motion in the ξ-*eta* plane is an ellipse, twice as long as it is deep. The period of this motion is the same as that of the orbit of the space station.

The out-of-plane motion is also a sinusoid, with the same period. Because it is independent of the ξ-η motion, its magnitude and the phase difference are arbitrary, depending only on the initial conditions ζ_0 and $\dot{\zeta}_0$.

Example 9.1. The space station is in orbit 600 km above the earth. It is desired to put a satellite in orbit near the station, to keep watch for signs of damage. At its apogee, the satellite is at 610 km altitude, and the station is between it and the earth. Take this point to be the initial condition.

1. What is the value of $\dot{\eta}_0$ if the energy matching condition is met (recall that it must be met, or the satellite will drift away from the station over several orbits).
2. The out of plane motion is to have the same magnitude as the radial motion. What is the value of $\dot{\zeta}_0$?

Since the station is between the satellite and the earth at the initial condition, we have

$$\xi_0 = 10\,\text{km}; \qquad \eta_0 = \zeta_0 = 0.$$

The mean motion of the station orbit is

$$n = \sqrt{\mu_\oplus/a^3} = \sqrt{\frac{398601}{(6378+600)^3}} = 0.001083\,\text{s}^{-1}$$

From (9.24), we have

$$\dot{\eta}_0 = -2n\xi_0 = -0.02166\,\frac{\text{km}}{\text{s}}.$$

The out of plane motion is given by (9.17). Since $\zeta_0 = 0$, in this case we have

$$\zeta(t) = (\dot{\zeta}_0/n)\sin nt$$

which will be maximized at $\zeta = \dot{\zeta}_0/n$ when $\sin nt = 1$. Therefore,

$$\dot{\zeta}_0 = 10n \quad \Longrightarrow \quad \dot{\zeta}_0 = 0.01083\,\frac{\text{km}}{\text{s}}.$$

The relative motion of the satellite about the space station is shown in Figure 9.3. The radial vs. along track motion is precisely the two-by-one ellipse that is expected. The projection of the out-of-pane motion vs. along track motion appears as a line. This is an edge view of the actual elliptic motion. ♠

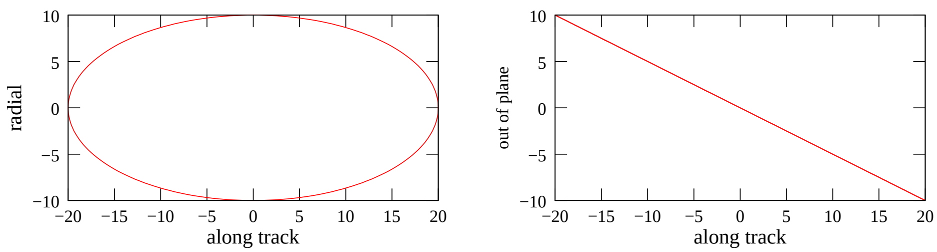


Figure 9.3: Relative motion of satellite (example 9.1).

9.1.4 The Transition Matrix

The natural motion of the craft relative to the space station is fully defined by the six initial conditions. Re-writing (9.17), (9.22), and (9.23) to separate the effect of each initial value, and writing the result in matrix-vector form, provides

$$
\begin{aligned}
\left\{\begin{array}{c} \xi(t) \\ \eta(t) \\ \zeta(t) \end{array}\right\} &= \left[\begin{array}{ccc} 4-3\cos nt & 0 & 0 \\ 6\sin nt - 6nt & 1 & 0 \\ 0 & 0 & \cos nt \end{array}\right] \left\{\begin{array}{c} \xi_0 \\ \eta_0 \\ \zeta_0 \end{array}\right\} \\
&+ \frac{1}{n}\left[\begin{array}{ccc} \sin nt & 2-2\cos nt & 0 \\ 2\cos nt - 2 & 4\sin nt - 3nt & 0 \\ 0 & 0 & \sin nt \end{array}\right] \left\{\begin{array}{c} \dot{\xi}_0 \\ \dot{\eta}_0 \\ \dot{\zeta}_0 \end{array}\right\}. \qquad (9.29)
\end{aligned}
$$

The derivatives of the motion are written easily enough by taking the derivative of (9.29) with respect to time. Assembling the results as above gives

$$
\begin{aligned}
\left\{\begin{array}{c} \dot{\xi}(t) \\ \dot{\eta}(t) \\ \dot{\zeta}(t) \end{array}\right\} &= \left[\begin{array}{ccc} 3n\sin nt & 0 & 0 \\ 6n\left(\cos nt - 1\right) & 0 & 0 \\ 0 & 0 & -n\sin nt \end{array}\right] \left\{\begin{array}{c} \xi_0 \\ \eta_0 \\ \zeta_0 \end{array}\right\} \\
&+ \left[\begin{array}{ccc} \cos nt & 2\sin nt & 0 \\ -2\sin nt & 4\cos nt - 3 & 0 \\ 0 & 0 & \cos nt \end{array}\right] \left\{\begin{array}{c} \dot{\xi}_0 \\ \dot{\eta}_0 \\ \dot{\zeta}_0 \end{array}\right\}. \qquad (9.30)
\end{aligned}
$$

These two equations are now combined into one.

$$
\left\{\begin{array}{c} \boldsymbol{\rho}(t) \\ \dot{\boldsymbol{\rho}}(t) \end{array}\right\} = \left[\begin{array}{cc} \boldsymbol{\Phi}_{11}(t) & \boldsymbol{\Phi}_{12}(t) \\ \boldsymbol{\Phi}_{21}(t) & \boldsymbol{\Phi}_{22}(t) \end{array}\right] \left\{\begin{array}{c} \boldsymbol{\rho}_0 \\ \dot{\boldsymbol{\rho}}_0 \end{array}\right\} \qquad (9.31)
$$

The matrix in (9.31) is called a *transition matrix*. Note that any arbitrary time t_1 can be taken to be the initial time. Then the state at any time t_2 is written

$$\left\{ \begin{array}{c} \boldsymbol{\rho}(t_2) \\ \dot{\boldsymbol{\rho}}(_2) \end{array} \right\} = \left[\begin{array}{cc} \boldsymbol{\Phi}_{11}(t_2 - t_1) & \boldsymbol{\Phi}_{12}(t_2 - t_1) \\ \boldsymbol{\Phi}_{21}(t_2 - t_1) & \boldsymbol{\Phi}_{22}(t_2 - t_1) \end{array} \right] \left\{ \begin{array}{c} \boldsymbol{\rho}(t_1) \\ \dot{\boldsymbol{\rho}}(t_1) \end{array} \right\} \tag{9.32}$$

A set of canonical units for this problem can be derived based on the nominal orbit radius r_s. Since $\mu = 1$ in canonical units, this results in $n = 1$ and the period P_s of the reference orbit as 2π time units. Such a unit set allows the time to be specified easily in terms of fractions of an orbit.

9.1.5 Impulsive Rendezvous

The transition matrix of the last section can be used to solve for impulsive rendezvous in circular orbit.

Example 9.2. An astronaut is ten kilometers above and ten kilometers behind his spacecraft, which is on a circular orbit of radius 10000 kilometers about the earth. If his instantaneous velocity relative to the craft is 10 meters per second in the local x direction, what are the thrusts required to return to the craft, arriving when $nt = \pi/2$?

The problem can be solved in either canonical or standard units. Choosing standard units, we have

$$n = \sqrt{\mu_\oplus / r_s^3} = 6.3135 \times 10^{-04}\,\mathrm{s}^{-1}.$$

Note that the time appears in (9.29) and (9.30) only multiplied by n, so the actual value of the final time is not needed.

The initial conditions are

$$\boldsymbol{\rho}_0 = \left\{ \begin{array}{c} 0 \\ -10 \\ 10 \end{array} \right\} \mathrm{km} \qquad \text{and} \qquad \dot{\boldsymbol{\rho}}_0 = \left\{ \begin{array}{c} 10 \\ 0 \\ 0 \end{array} \right\} \frac{\mathrm{km}}{\mathrm{s}}.$$

These conditions will not result in the desired rendezvous, so the problem is to compute the relative velocity after the impulse so that rendezvous will happen. Recalling that an impulse does not change the position, only the velocity, we use (9.29) to solve as

$$\dot{\boldsymbol{\rho}}_0^+ = \boldsymbol{\Phi}_{12}^{-1}(\pi/2)\left[\boldsymbol{\rho}(\pi/2) - \boldsymbol{\Phi}_{11}(\pi/2)\boldsymbol{\rho}_0\right] \tag{9.33}$$

where as usual the superscript + denotes the velocity immediately after the impulse.

Now, the desired final position is $\boldsymbol{\rho}(\pi/2) = 0$. Inserting this, along with the initial position, into (9.33) and solving produces

$$\dot{\boldsymbol{\rho}}_0^+ = \left\{ \begin{array}{c} -3.841 \times 10^{-03} \\ 1.920 \times 10^{-03} \\ 0 \end{array} \right\} \frac{\mathrm{km}}{\mathrm{s}} = \left\{ \begin{array}{c} -3.841 \\ 1.920 \\ 0 \end{array} \right\} \frac{\mathrm{m}}{\mathrm{s}}.$$

The impulse needed is this value minus that before the impulse. Thus,

$$\Delta v_1 = \dot{\boldsymbol{\rho}}_0^+ - \dot{\boldsymbol{\rho}}_0 = -13.84\hat{\imath} + 1.92\hat{\jmath} \text{ m/s}.$$

We now use (9.30) to compute the velocity of the astronaut when he arrives at the spacecraft. Since he will want a velocity of zero relative to the craft, he will use an impulse to negate the arrival velocity. Therefore,

$$\Delta v_2 = -\left[\boldsymbol{\Phi}_{21}\boldsymbol{\rho}_0 + \boldsymbol{\Phi}_{22}\dot{\boldsymbol{\rho}}_0^+\right]. \tag{9.34}$$

Solving this produces

$$\Delta v_2 = -3.841\hat{\imath} - 1.920\hat{\jmath} + 6.314\hat{k} \text{ m/s}.$$

The relative motion is shown in Figure 9.4. ♠

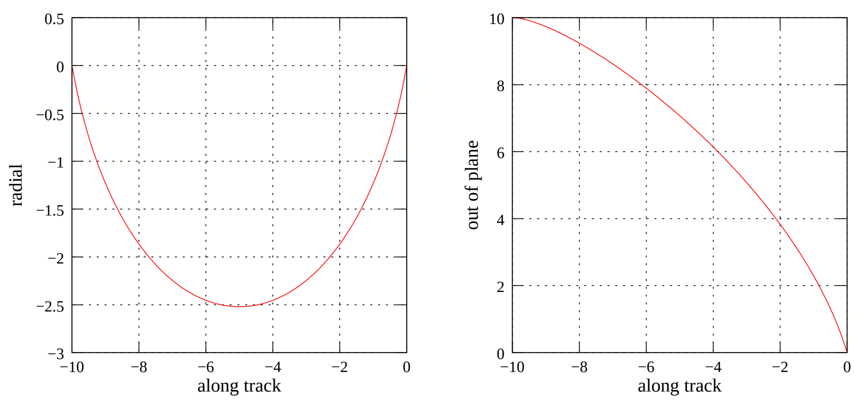


Figure 9.4: Motion of astronaut relative to spacecraft (example 9.2).

Equations (9.33) and (9.34) are valid in general for impulsive rendezvous problems in the Clohessy-Wiltshire formulation. However, they work only if the submatrix $\boldsymbol{\Phi}_{12}$ is invertible. It almost always is, but there are certain values of nt for which it is not.

In the example, the astronaut was heading to the origin of the local system. The only effect this has on the problem is to set the desired final value of $\boldsymbol{\rho}$ to zero.

As a final note of caution, it should be remembered that the Clohessy-Wiltshire equations are *approximations*. They are very close to reality for motion that is quite close to the reference circular orbit. As the motion gets larger, the accuracy decays. A shortcoming of this approach is that it does not provide a way to estimate the size of the error.

9.2 Natural Motion and Orbital Elements

In the last section, the differential equations for the motion of a spacecraft relative to another body were derived. Those equations are valid in general, but could only be solved if the thrust is known. It was a fairly straightforward process, however, to solve for the natural relative motion, when the reference orbit was circular.

The equations for relative motion can be derived as in section 9.1 for non-circular reference orbits as well, but the results are not so easily solved. As interest in satellite clusters increased, many investigators addressed these problems, and transition matrices for elliptic orbits were derived. In this section, we take a different approach, expanding the orbital elements directly. This will allow an examination of the qualitative motion of the spacecraft relative to the station.

9.2.1 Expansions of the Orbital Elements

In general, the idea of series expansion is that a quantity is expanded *about* a nominal value of that quantity, in terms of some (usually small) parameter. In the last section, the equations of motion were expanded by taking a Taylor expansion of the equations of motion. The nominal value of the motion was the circular reference orbit. The expansion parameter was the vector joining the body of interest to the origin of the local frame; this was small relative to the radius of the reference orbit.

In this section, the motion of a body in orbit is expanded about the mean anomaly, and the expansion parameter will be the orbital eccentricity. The techique used is the Lagrange method, whic is presented in Appendix C.2. It is not the only technique available, but it works well in expanding orbital elements, because it is well-suited to relations of the form of the eccentric anomaly.

The eccentric anomaly is expanded in the appendix, as an example of the method. Using the same technique to expand $\cos E$ provides

$$\cos E = \cos M - \frac{e}{2}(1 - \cos 2M) + \frac{3e^2}{8}(\cos 3M - \cos M) + \cdots \tag{9.35}$$

This can then be substituted into equation (3.11)

$$\frac{r}{a} = 1 - e\cos E$$

to get

$$\frac{r}{a} = 1 - e\cos M + \frac{e^2}{2}(1 - \cos 2M) - \frac{3e^3}{8}(\cos 3M - \cos M) + \cdots \tag{9.36}$$

Clearly, if e is small, the terms in the expansion become small very quickly. The series is valid for fairly large values of eccentricity, as well, though it fails to converge for some values of M when e is greater than about $e = 0.663$.

Finding an expansion for the true anomaly is somewhat more difficult, and the exercise does not provide increased understanding of the result,[5] and so will be neglected here. In the end, the result is the expansion

$$\nu = M + 2e\sin M + \frac{5e^2}{4}\sin 2M + \frac{e^3}{12}\left(13\sin 3M - 3\sin M\right) + \cdots \tag{9.37}$$

Note that these expansions are for the motion of an individual spacecraft; nothing has been said about a second vehicle. However, as was noted in section 9.1, the only way for two vehicles in natural motion to remain near each other is if their orbital energies match. If the energies match, it follows as well that the semimajor axes and the mean motions match. In such a case, the mean anomaly of one vehicle minus that of the other will be a constant.

9.2.2 An Order Analysis

Consider the motion of one satellite relative to another, when the second satellite, which we take as the reference, is in an orbit that is significantly elliptic. If the second satellite is to be close to the reference at all points on the orbit, they must have very similar arguments of periapsis. Since the mean anomaly is measured from periapsis, it follows that

$$M = M_0 + \delta M, \tag{9.38}$$

where M is the mean anomaly of the satellite in which we are interested, M_0 is the mean anomaly of the reference satelite, the difference δM is small. Applying (9.37) to each satellite and taking the difference, we have

$$\begin{aligned} \nu - \nu_0 \;=\;& \delta M + 2(e_0+\delta e)\sin(M_0+\delta M) + \frac{5}{4}\left(e_0+\delta e\right)^2\sin\left(2M_0+2\delta M\right) \\ & -2e_0\sin M_0 - \frac{5}{4}e_0^2\sin M_0 + \cdots \end{aligned} \tag{9.39}$$

Invoking some well-known trigonometric identities provides

$$\begin{aligned} \nu - \nu_0 \;=\;& \delta M \\ & +2e_0\left(\sin M_0\cos\delta M + \cos M_0\sin\delta M\right) - 2e_0\sin M_0 \\ & \quad +2\delta e\left(\sin M_0\cos\delta M + \cos M_0\sin\delta M\right) \\ & +\frac{5}{4}e_0^2\left(\sin 2M_0\cos 2\delta M + \cos 2M_0\sin 2\delta M\right) \\ & \quad +\frac{5}{2}e_0\delta e\left(\sin 2M_0\cos 2\delta M + \cos 2M_0\sin 2\delta M\right) \\ & +\frac{5}{4}\delta e^2\left(\sin 2M_0\cos 2\delta M + \cos 2M_0\sin 2\delta M\right) - \frac{5}{4}e_0^2\sin 2M_0 + \cdots \end{aligned} \tag{9.40}$$

[5]Though it is enjoyable as a mathematical exercise.

To perform an order analysis means to identify terms that have the same order in small parameters. In the last section, we noted that the magnitude of $\boldsymbol{\rho}$ was very small relative to r_s. Any term that included higher powers of $\boldsymbol{\rho}$ were therefore considered negligible. In this case, things are not quite as simple, as there are more variables to consider.

Consider the along-track relative motion. The distance that the satellites are apart will be very nearly

$$\eta \approx r_0(\nu - \nu_0).$$

If this is considered small relative to r_0, it is clear that $\nu - \nu_0$ is the small term. In (9.39), δM appears by itself, so if $\nu - \nu_0$ is small, so must be δM.

Now think about the radial distance. Since the orbits must be similar, the satellites must have similar periapsis radii. Recall that

$$r_p = a(1 - e)$$

so that the difference in periapsis radii is

$$r_p - r_{p,0} = a\delta e.$$

Since a is not small, we have that δe must be small. We now have two small parameters.

Note: It is possible one of δe or δM is zero, or at least very small relative to the other. However, the important thing for our analysis is that neither is *larger* than "small". ♣

Consider the second term on the right hand side of (9.40). Because δM is a small parameter, it is necessary to expand the trigonometric terms containing it. Recalling that

$$\begin{aligned}
\sin\alpha &= \alpha - \frac{\alpha^3}{6} + \frac{\alpha^5}{5!} + \cdots \\
\cos\alpha &= 1 - \frac{\alpha^2}{2} + \frac{\alpha^4}{4!} + \cdots
\end{aligned}$$

shows that sine terms will contribute a first-order term, while cosine terms contribute a zeroth- and a second-order term. Both also contribute higher-order terms, but for now it is sufficient to consider only orders zero through two. Expanding the second term through second order in δM (there are no terms containing δe) provides.

$$2e_0\left(\sin M_0 \cos\delta M + \cos M_0 \sin\delta M\right) = 2e_0 \sin M_0\left(1 - \delta M^2/2\right) + 2e_0\cos M_0\delta M \tag{9.41}$$

The term $2e_0 \sin M_0$ on the right hand side will be subtracted out by the third term in (9.40), eliminating the zeroth-order term. The only remaining term of first-order is $2e_0 \cos M_0\delta M$.

Chapter 10

Rocket Performance

The only feasible method so far devised for achieving the velocities necessary to orbit satellites, or to send objects (and people) beyond the gravity well of the Earth, is the chemical rocket. Many other schemes have been proposed, most impossible, but none have so far been implemented.

Some of these other schemes have to their credit great imagination, and some the possibility of great achievements. Several authors over the last few centuries have proposed flying to the moon in balloons, or in vehicles harnessed to birds. Jules Verne's idea of using a mammoth cannon to fire a shell to the moon is also arresting, and equally impossible (at least with any chance of survival for passengers).

To get goods from the moon to the Earth, the lunar inhabitants in Robert Heinlein's *The Moon is a Harsh Mistress* use a catapult. This is more feasible, since the gravity of the moon is much less than that of the Earth. And once in space, there are proposals for solar sails, which are basically huge mirrors, that use the very small momentum of light to accelerate arbitrarily large bodies to interstellar speeds.

All of these ideas and more are beyond the scope of this text. In this chapter, we will restrict ourselves to a basic discussion of the chemical rocket. We will first derive the most basic equations for describing rocket performance outside of the atmosphere and gravity of a planet, and then discuss the forces due to these factors. We will look at the sounding rocket. Finally, we will discuss single- and multi-stage rocket performance in terms of mass fractions, the non-dimensional factors often used to describe vehicles.

10.1 Thrust of a Rocket

For our purposes, a rocket will be any device that creates thrust by expelling only stored mass. This definition excludes jet engines, which primarily use their stored fuel to increase the energy of the air stream that flows through them. The only such rockets that achieve sufficient thrust to enable launch from the surface of the

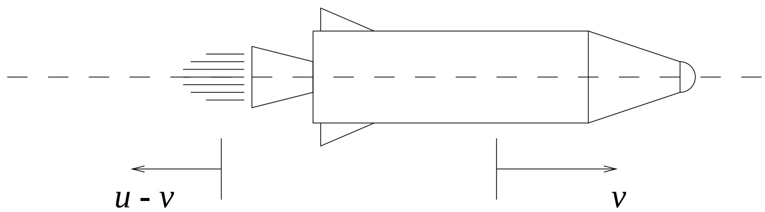


Figure 10.1: Sketch of rocket and exhaust

Earth use chemical fuel, though electrical thrusters have been developed for use in low-thrust applications once in orbit.

A chemical rocket works by burning stored fuel (typically referred to as propellant) and allowing the exhaust gases to escape through a nozzle at the back. These gases are flowing at a velocity u relative to the rocket, which itself is traveling at a velocity v. To compute the resulting thrust, consider the rocket at time t, and at some later time $t + \Delta t$.

At time t, the rocket has mass m, and is moving at some velocity v. After the interval Δt, the mass has decreased because some Δm of propellant has been expelled, and the velocity has increased by some increment Δv. Thus the momentum at times t and $t + \Delta t$ are

$$M_t = mv \tag{10.1}$$

$$M_{t+\Delta t} = (m - \Delta m_p)(v + \Delta v) + \Delta m(v - u) \tag{10.2}$$

Since momentum is conserved, these two can be set equal to each other to get

$$\begin{aligned} mv &= (m - \Delta m_p)(v + \Delta v) \\ &\implies \quad m\Delta v = u\Delta m_p - \Delta v \Delta m_p \end{aligned} \tag{10.3}$$

Dividing through by Δt and taking the limit as $\Delta t \to 0$ gives

$$m\dot{v} = u\dot{m}_p \tag{10.4}$$

The term on the left is the instantaneous change in the momentum of the rocket, and thus is equal to the thrust imposed upon it by the expulsion of propellant. We thus have that

$$T = u\dot{m}_p \tag{10.5}$$

This equation tells us that the thrust of a rocket is dependent both upon how much mass flows through the nozzle, and on the speed it attains relative to the rocket as it leaves. The two obvious ways to improve the performance of a rocket are then to increase the mass flow rate $\dot{m}_p$, or the exhaust velocity u. The first of these can be easily enough accomplished by building a bigger rocket motor, but this requires more structural mass and may defeat the purpose. The second requires using very high-energy fuels, or changing the basic technology of the rocket. Propellant chemists work at increasing the energy derived from chemical propellants. The goal of technologies such as nuclear and electrical rockets is to achieve exhaust velocities well above those of chemical rockets.

10.1.1 The Rocket Equation

The mass flow rate in eqn. (10.4) is positive because it refers to the mass of propellant. The mass of the rocket, which includes all unburnt propellant, changes as well, and the rate of change has the opposite sign. We can thus re-write (10.4) as

$$m\dot{v} = -u\dot{m}$$

We can solve this equation:

$$m\frac{dv}{dt} = -u\frac{dm}{dt} \quad \Longrightarrow \quad mdv = -udm \quad \Longrightarrow \quad dv = -u\frac{dm}{m}$$

We make the assumption that u is a constant (which is in fact quite a good assumption), which allows us to integrate:

$$\begin{aligned} \int_{v_0}^{v} dv &= -u \int_{m_0}^{m} \frac{dm}{m} \\ \Longrightarrow \quad v - v_0 &= u \ln(m_0/m) \end{aligned}$$

Replacing $v - v_0$ with the commonly used Δv gives us

$$\Delta v = u \ln(m_0/m) \tag{10.6}$$

This is one form of the *rocket equation*, which describes the ideal performance of a rocket. In various forms, this equation will be the main component of the rest of this chapter.

10.2 The Sounding Rocket

The simplest case of a rocket trajectory under gravity is the sounding rocket[1]. These rockets are launched vertically, usually carrying an instrument package that will then be allowed to come back to earth on a parachute.

In the ideal case, a sounding rocket will travel a straight line directly upwards. In reality, of course, winds, thrust misalignment, and asymmetries in the rocket and

[1] The term is an analogue of the naval term. "Sounding" refers to lowering a weight on a rope over the side of a sea-going vessel to determine the depth of the water.

payload will cause a small variation from this, but we will ignore these effects in this analysis.

The sounding trajectory will consist of two parts, the boost phase and the coast phase. During the first of these, the rocket motor is burning and the vehicle is accelerating against gravity. During the second, the motor has burned out, and the rocket will coast until the kinetic energy it had at the end of the boost is gone and the velocity has gone to zero.

10.2.1 Powered Boost

The boost phase of the trajectory will typically last several seconds. During this phase, the vehicle will gain velocity and altitude. The altitude gained will be sufficiently small that we may ignore the variation of gravitational attraction with altitude, and we will take gravity to be constant. We will also ignore any effects due to the atmosphere.

The equations of motion are therefore simple. The vertical acceleration is the difference between the effects of thrust and of gravity, and the velocity is the rate of change of altitude. Maintaining our assumption of constant thrust and exhaust velocity, we also have a constant mass flow rate from the motor, and the equations are

$$\dot{v} = -\frac{u\dot{m}}{m} - g \tag{10.7}$$

$$\dot{h} = v \tag{10.8}$$

$$\dot{m} = -k \tag{10.9}$$

Here, the constant k is the mass flow rate of the burned fuel from the motor[2].

We will begin by considering the final equation, (10.9). This equation can be written

$$dm = -kdt$$

which allows us to write the mass as a function of time as

$$m(t) = m_0 - kt \tag{10.10}$$

This will be useful in evaluating the other two expressions.

Consider eqn. (10.7). We can re-write this as

$$m\dot{v} = -u\dot{m} - mg$$

The first term on the right is simply the equation we derived at the beginning of the chapter, and thus gives us the rocket equation when we integrate. The second term is evaluated as

$$\dot{v} = -g \quad \Longrightarrow \quad v = -gt$$

[2] Liquid-fueled motors, or solid motors with varying burn rates, will not have a constant flow rate. Most sounding rockets are made with simple, inexpensive solid fuel motors, however.

(ignoring any initial velocity). Thus the equation as a whole gives

$$v = v(0) + u\ln(m_0/m) - gt \tag{10.11}$$

This is correct, but the use of both t and m as independent variables on the right hand side is needlessly confusing. Instead, we use eqn. (10.10) to write time as

$$t = \frac{m_0 - m}{k}$$

Now we can write the velocity as a function of either mass or time as

$$\begin{aligned} v(t) &= v(0) + u\ln\left(\frac{m_0}{m_0 - kt}\right) - gt && (10.12)\\ v(m) &= v(0) + u\ln(m_0/m) - g\frac{m_0 - m}{k} && (10.13) \end{aligned}$$

Note that, until we wrote mass as a function of time, we had made no assumptions of constant flow rate in deriving the velocity. Thus eqn. (10.11) is valid so long as the exhaust velocity is constant, while the last two are valid only if the mass flow rate is also constant.

Solving for the altitude is somewhat more complicated. We can again choose to solve for a function of mass or time, and we will choose to solve for mass. Thus we re-write eqn. (10.8) using (10.13) as

$$dh = v(0)dt + u\ln(m_0/m)dt - g\frac{m_0 - m}{k}dt$$

Then using

$$t = \frac{m_0 - m}{k} \qquad \Longrightarrow \qquad dt = -\frac{dm}{k}$$

we get

$$dh = -\frac{v(0)dm}{k} + \frac{u}{k}\ln\left(\frac{m}{m_0}\right)dm - \frac{g}{k}\frac{m - m_0}{k}dm \tag{10.14}$$

We integrate each term on the right hand side of (10.14) separately. Up to this point, we have not worried about the notation of integrating a function of mass using limits of integration that are themselves functions of mass. To make it clear that the values used as limits of integration are specific values, rather than functions, we will use $\bar{m}$ as the limit of integration, and replace it with m at whatever point is convenient when we are finished.

The first term integrates simply to

$$\int_{m_0}^{\bar{m}} -\frac{v(0)dm}{k} = \frac{(m_0 - \bar{m})v(0)}{k}$$

The second we integrate[3] as

$$\frac{u}{k}\int_{m_0}^{\bar{m}} \ln\left(\frac{m}{m_0}\right) dm = \frac{u}{k}\left[\bar{m}\ln\left(\frac{\bar{m}}{m_0}\right) + m_0 - \bar{m}\right]$$

The third becomes

$$\frac{g}{k^2}\int_{m_0}^{\bar{m}} (m - m_0)dm = \frac{g(\bar{m} - m_0)^2}{2k^2}$$

The altitude as a function of mass is then

$$h = \frac{(m_0 - m)v(0)}{k} + \frac{u}{k}\left[m\ln\left(\frac{m}{m_0}\right) + m_0 - m\right] - \frac{g(m - m_0)^2}{2k^2} \tag{10.15}$$

In the usual case, the initial velocity will be zero. We can then find the velocity when the motor burns out from

$$h_{bo} = \frac{u}{k}\left[m_{bo}\ln\left(\frac{m_{bo}}{m_0}\right) + m_0 - m_{bo}\right] - \frac{g(m_{bo} - m_0)^2}{2k^2} \tag{10.16}$$

10.2.2 Coast Phase

After the motor burns out, the rocket will continue to coast upwards until all of the kinetic energy it attained during the boost phase is converted to potential energy. In this phase, the rocket will cover sufficient altitude that the variation in gravity has a significant effect. Therefore, we replace our constant gravity with a function of altitude:

$$g(h) = g_0\frac{R^2}{(R+h)^2} \tag{10.17}$$

In this equation, R is the radius of the Earth (or the moon, or whatever body the rocket is being launched from), h is the altitude above the surface, and g_0 is the gravitational acceleration at the surface.

Since the gravitational force field is conservative, we can equate the kinetic energy of the rocket at burnout to the work done by gravity during the coast. We then have

$$\begin{aligned} \frac{m_{bo}v_{bo}^2}{2} &= m_{bo}g_0\int_{h_{bo}}^{h_{bo}+h_c} \frac{R^2 dh}{(R+h)^2} \\ \implies \quad \frac{v_{bo}^2}{2} &= \frac{g_0R^2}{R+h_{bo}} - \frac{g_0R^2}{R+h_{bo}+h_c} \end{aligned}$$

which, after some simplification, gives the coasting altitude to be

$$h_c = \frac{v_{bo}^2(R+h_{bo})^2}{2g_0R^2 - v_{bo}^2(R+h_{bo})} \tag{10.18}$$

Adding this altitude to the altitude at burnout will give the final altitude of the sounding rocket.

[3]We find in tables that $\int \ln x\, dx = x\ln x - x$.

Example 10.1. Given a sounding rocket with a payload of 12 pounds,[4] 8 pounds of structure, and 12 pounds of propellant with an exhaust velocity of 10000 feet per second, what is the final altitude achieved if the burn time is 12 seconds? How does it change if the burn time is only 6 seconds?

Solution: The equations we just derived give us

$$\begin{aligned} v(t_{bo}) &= u\ln(m_0/m_{bo}) - gt \\ \implies \quad v_{bo} &= 10000\ln[(12+8+12)/(12+8)] - 12\cdot 32.174 \\ &= 4700 - 386.1 = 4314 \text{ ft/s} \end{aligned}$$

Note that we went ahead and used pounds as a unit of mass here. We can get away with this because we are using it in a ratio, where the conversion factor will cancel out. This is in general dangerous – the notion of "12 pounds of propellant" is nonsensical anywhere but on the surface of the Earth.

The equation for burnout altitude also has mass only used in ratios, since the masses inside the brackets in eqn. (10.16) are divided by the mass flow rate outside. Thus, so long as we are careful to use consistent units, we can set $k = 12$lb-m/12 seconds and compute

$$h_{bo} = \frac{10000}{1}\left[20\ln(20/32) + 32 - 20\right] - \frac{32.174}{2}(20-32)^2 = 2.370\times 10^4 \text{ ft}$$

The additional altitude due to the coast phase is given by eqn. (10.18) to be (using $R = 2.0925673\times 10^7$ feet)

$$h_c = \frac{4314^2\cdot 20949355^2}{2\cdot 32.174R^2 - 4314^2(20949355)} = 2.939\times 10^5 \text{ ft}$$

The total altitude of the rocket thus comes out to be 317,600 feet.

If the burn time is shortened to 6 seconds, k becomes 2 lb-m per second, and the numbers become

$$\begin{aligned} v_{bo} &= 4507 \text{ ft/s} \\ h_{bo} &= 1.286\times 10^4 \text{ ft} \\ h_c &= 3.209\times 10^5 \text{ ft} \end{aligned}$$

for a total altitude of 333,800 feet. ♠

The results of this example help to explain the optimal use of propellant in sounding rockets. In is well-known to rocket designers that the most efficient use of propellant when thrusting against gravity is to burn as quickly as possible. It is clear

[4] When the term "pound" is used as a unit of mass, it is understood that the unit is the lb-mass, defined as the mass in slugs times g_c (giving us the amount of mass that weighs n pounds under normal gravity). While this is not a generally accepted unit of mass, it is still often used.

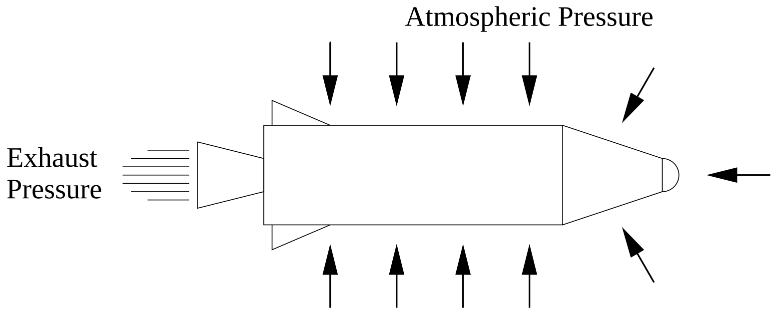


Figure 10.2: Atmospheric pressure on a rocket

from eqn. (10.12) that this results in the highest burnout velocity. It is not as clear that the reduced altitude at burnout is more than recovered during the coast.

Obviously, the limiting case of the impulsive burn is not possible, as it results in infinite acceleration. The structural limits of the rocket and payload often limit the acceleration, and in this way define the burn time. Even when this is not a problem, however, the "faster-is-better" rule may not hold when the effects of atmospheric drag are included in the equations.

10.3 Aerodynamic Effects on Rocket Performance

While a key use of rockets, and the use that makes them of interest to us, is their ability to function in space, the majority of rockets actually spend most or all of their working lives in the atmosphere. Missiles, for example, are commonly powered by solid rocket motors. The sounding rocket also generally does all of its thrusting while still well in the dense lower atmosphere.

10.3.1 Pressure Thrust

The two primary effects of the atmosphere are drag and an additional term in the thrust equation. We will look at the second of these first.

Consider a rocket in the atmosphere. The arrows in Figure 10.2 represent the atmospheric pressure, which is evenly distributed around the body and acts normal to the surface everywhere, with the exception of the *exit plane*. The exit plane is the area of the rocket nozzle at the end, and the pressure acting on the rocket at this point is the exhaust pressure P_e.

Summing the pressure forces about the rocket body, we see that all of the forces other than those involving the exhaust pressure cancel themselves out, and we are left with a term in the direction of rocket travel that is due to the difference in exhaust pressure and atmospheric pressure:

$$T_{ex} = (P_e - P_{atm})A_e \tag{10.19}$$

That is, thrust due to the atmosphere is equal to the difference in exhaust and ambient pressure, times the area of the exit plane.

The pressure in the atmosphere varies with altitude. To first order, a good approximation is the *exponential atmosphere*, in which the pressure is given by

$$P_{atm} = P_0 e^{-h/\beta_P} \tag{10.20}$$

Here, P_0 is the pressure at sea level, and β_P is a proportionality constant. If h is measured in feet, the value of β_P is usually taken to be 23200. Because of this reduction in pressure with increasing altitude, the amount of thrust due to exhaust pressure increases as the rocket gains altitude.

The exhaust pressure is determined by two primary factors. One is the pressure in the combustion chamber of the rocket. Most chemical propellants burn faster at higher pressures. Also, the density of the exhaust gas is higher at higher pressure, so that more mass can flow through a smaller nozzle. For these reasons, small high-performance rocket motors often will have high combustion chamber pressures.

The exhaust pressure is given by the chamber pressure times the second major factor, which is determined by the *expansion ratio* of the nozzle. The expansion ratio is simply the exit area of the nozzle divided by the cross-sectional area at the smallest part, or *throat*, of the nozzle.

While it might seem that the ideal rocket would have a very high pressure thrust term, this is not the case. The exhaust pressure decreases as the exhaust gases flow through the expanding nozzle, so that a high pressure also tends to mean small area. More importantly, the gases attain higher velocity as they expand (an explanation of this is far beyond the scope of this book), and it is this velocity that gives the majority of the thrust. It turns out that the ideal case, yielding maximum thrust, is when the exhaust pressure exactly matches the ambient pressure.

Due to design constraints and the change of pressure with altitude, matching the pressures is seldom possible. For air-to-air missiles, which must fit under the wing of an aircraft, the nozzle size is limited and the performance must be very high, so the exhaust pressure can be in excess of 100 psi. For the space shuttle boosters, however, the chamber pressure is much lower, and the nozzles can be much larger. For this class of vehicle, the pressure term may actually be negative at launch, and become positive only as the atmospheric pressure gets low.

10.3.2 Drag

Aerodynamic drag is the force impeding the motion of a body through the atmosphere. The effect is proportional to the square of the velocity, so that while it is experienced by any body moving through the air, it is a far greater effect for high-speed vehicles such as aircraft and rockets.

There are several components of drag, from the friction of the air rubbing over the body to the energy required to push the air out of the path of the body. It is beyond our needs to investigate these thoroughly. Instead, we will simply note that drag on a rocket is usually characterized by the equation

$$D = \frac{1}{2}\rho v^2 S_{\text{ref}} C_D \tag{10.21}$$

In eqn. (10.21), ρ is the density of the atmosphere, and v is the velocity of the rocket relative to the atmosphere. The term S_{ref} (often simply S) is the *reference area*, which for a rocket is usually defined to be the cross-sectional area of the main rocket body, looking down the longitudinal axis. Since a rocket typically has a circular cross-section, this is simply $\pi d^2/4$, where d is the diameter of the rocket.

The term C_D is called the *Drag coefficient*. It is defined for a particular body through eqn. (10.21). That is, the drag is measured, and divided by $\rho v^2 S_{\text{ref}}/2$ to get the coefficient. The value of C_D varies with velocity (more precisely, with Mach number), but is nearly constant for most rockets above three times the speed of sound, where the vehicle spends most of its operating lifetime. Thus, for initial design work, C_D is usually taken to be a constant.

For typical rockets, the value of C_D at high Mach numbers will be in the range 0.12 to 0.16. The lower value will be for a very clean, sleek design.

If we include drag in the differential equation of motion, we have before burnout

$$\dot{v} = -\frac{1}{m}\left(u\dot{m} + \frac{1}{2}\rho v^2 S_{\text{ref}} C_D\right)$$

(recall that $\dot{m}$ is negative). After burnout, the thrust term of course goes to zero. This is a non-linear equation, and no closed-form solution exists for it. Instead, it is solved numerically on a case-by-case basis.

Example 10.2. Consider the sounding rocket of example 10.1. This rocket might typically have a diameter of four inches, and a drag coefficient of 0.14. We will assume the exponential atmosphere, which defines the density as a function of altitude to be

$$\rho(h) = \rho_0 e^{-h/\beta_\rho}$$

where the proportionality constant $\beta_\rho = 23800$ feet.

While this situation does not lead to analytical solutions, we can set up the equations of motion and integrate them numerically. For this example, we will ignore the thrust due to exhaust pressure, but include the gravity term. The equations of motion are

$$
\begin{array}{llll}
0 \le t \le t_{bo}: & & t > t_{bo}: & \\
& \dot{m} = -k & & \dot{m} = 0 \\
& \dot{v} = (uk - D)/m - g & & \dot{v} = -D/m - g \\
& \dot{h} = v & & \dot{h} = v
\end{array}
$$

Using 12 seconds as the burn time, we have a maximum drag of 79.4 pounds at burnout, which is the point of maximum velocity. The thrust is $T = uk = 10000(1/g_c) = 311$ pounds, so the drag is fully one-fourth of the thrust.

The burnout velocity is reduced by the drag to 3686 ft/s, and the altitude to 21650 ft. The final altitude is drastically reduced, with the vehicle attaining a maximum of only $h_f = 1.673 \times 10^5$ feet. ♠

Note that in this example, the mass and mass flow rate cannot be in pounds-mass, because the term D/m would not come out correctly. In this case, a consistent set of units *must* be used.

The problem of attaining the maximum altitude with a sounding rocket is known as the *Goddard problem*, after Robert Goddard, an American pioneer of rocketry. Too high an acceleration creates high velocities while still in the lower atmosphere, where the relatively great atmospheric density leads to large drag forces. Too low an acceleration, however, allows gravity too much time to act against the rocket. The mathematics of the tradeoff are very complicated. The solutions are unique to each rocket, and generally require a variable-thrust motor to implement.

10.4 Specific Impulse

Rocket fuels are often compared based on a number called the *Specific Impulse*. The specific impulse of a particular fuel is defined as the thrust obtained by burning a pound of fuel, multiplied by the amount of time required to burn the fuel. The higher this number, the better the fuel.

Under ideal circumstances, the thrust is given as seen in section 10.1 by the equation

$$T = u\dot{m}_p$$

Thus we can relate the specific impulse to the exhaust velocity as

$$I_{\text{sp}} = \int_0^t T dt = \int_0^t u \frac{dm_p}{dt} dt$$

Switching the variable of integration from time to mass, we note that one pound of fuel is $1/g_c$ slugs, so we get

$$\int_0^t u \frac{dm_p}{dt} dt = u \int_0^{1/g_c} dm = u/g_c$$

We therefore have the result that

$$u = g_c I_{\text{sp}} \tag{10.22}$$

Plugging this into eqn. (10.6), we get a second version of the rocket equation,

$$\Delta v = g_c I_{\text{sp}} \ln(m_0/m) \tag{10.23}$$

The unit of I_{sp} is seconds, and thus eqn. (10.23) works in all unit systems in which the time is measured in seconds. In the metric system, for instance, $g_c = 9.80665$ m/s^2 and the exhaust velocity will be in meters per second.

As was seen in section 10.3, the actual thrust seen by a rocket will depend on several factors beyond the chemical content of the fuel itself. In order to make eqn. (10.23) meaningful, these factors must in some way be accounted for. This is typically done by using a value for specific impulse that is a bit lower than the ideal.

The ideal specific impulse is that which multiplies the burn time by the highest possible thrust. This thrust is that attained if the motor is burned in a vacuum (so that P_{atm} is zero) and the exhaust nozzle allows the exhaust gas to fully expand. This value is often known as the *vacuum specific impulse*, and is thus denoted I_{sp_v}.

A value often used in preliminary design is the *delivered specific impulse*. Since the delivered thrust depends on the circumstances under which the fuel is burned, this value can change quite a bit. The rocket designer can deduce a good value by starting with the vacuum specific impulse and accounting for the other effects. She then has a value of I_{sp_d} which can be used in further preliminary design and analysis work.

As seen is section 10.1, the thrust developed by a rocket motor is determined in the main by the velocity of the exhaust gases. For a high velocity, the amount of energy must be high, implying a high combustion temperature, and the molecular weight of the gases must be low. One of the highest possible values of specific impulse is attained by burning a combination of liquid hydrogen and liquid oxygen, which results in $I_{\mathrm{sp}} = 457$ s. The space shuttle main engine achieves a delivered I_{sp} of 455 s, for an efficiency of over 99.5%.

10.5 Single Stage Performance and Mass Fractions

We saw in our earlier derivation of the rocket equation that the velocity capacity of a rocket is determined by the fraction m_0/m_{bo}. This value is known as the *mass ratio*. This ratio, along with the specific impulse of the motor, defines the performance of the vehicle, at least when the effects of atmosphere and gravity are ignored. Since many of these effects can be taken into account by judiciously altering the value of I_{sp} assumed delivered, this is actually a good approximation, and gives an excellent first approximation of the requirements and performance of the rocket engines needed for a mission.

The total mass of a single stage rocket vehicle is divided into three categories: the payload mass P, the propellant mass m_p, and the structural mass m_s. Of these, only the payload is "profitable" mass; the structural mass is dead weight that must be carried along to house the propellant and provide it a place in which to burn and a nozzle through which to exit.

The overall mass fraction of the vehicle is determined in part by the technology of the rocket motor, and in part by the payload. The motor technology can be described separately from the payload by the *structural factor* β,[5] defined as the structural mass over the mass of structure plus the mass of propellant:

$$\beta = \frac{m_s}{m_s + m_p} \tag{10.24}$$

We add to the motor the mass of the payload, giving us the total mass of the vehicle, and the part of that mass which is useful payload. The fraction of initial mass that consists of the payload is the *payload mass fraction* (or *payload fraction* or *payload ratio*),

$$\pi = \frac{P}{m_0} = \frac{P}{P + m_s + m_p} \tag{10.25}$$

These are the two mass fractions we will use to investigate vehicle performance. Others can be defined; in particular, the *propellant mass fraction* is often used. It is usually defined as the mass of propellant divided by the initial mass of the vehicle.

The equation used to estimate the performance is the rocket equation, which we recall is

$$\Delta v = g_c I_{\rm sp} \ln(m_0/m_{bo})$$

We write this in terms of the structural factor and the payload fraction:

$$\begin{aligned}\frac{m_{bo}}{m_0} &= \frac{m_s + P}{m_s + m_p + P} = \frac{m_s + P + (m_p - m_p)}{m_s + m_p + P} \\ &= 1 - \frac{m_p}{m_s + m_p + P} = 1 - \frac{m_p}{m_0} = 1 - \frac{m_s + m_p}{m_0}\frac{m_p}{m_s + m_p}\end{aligned}$$

Now,

$$\frac{m_s + m_p}{m_0} = \frac{m_0 - P}{m_0} = 1 - \pi$$

and

$$\frac{m_p}{m_s + m_p} = \frac{m_p + m_s - m_s}{m_s + m_p} = 1 - \beta$$

so that

$$\frac{m_{bo}}{m_0} = 1 - (1 - \pi)(1 - \beta) = \beta + (1 - \beta)\pi$$

Finally, we have

$$\Delta v = g_c I_{\rm sp} \ln(m_0/m_{bo}) = -g_c I_{\rm sp} \ln[\beta + (1 - \beta)\pi] \tag{10.26}$$

Example 10.3. Suppose we have a structural factor of $\beta = 0.08$ and a delivered specific impulse of $I_{\rm sp} = 230$. What is the maximum velocity (assuming a start from zero) that can be attained?

[5] Various authors use different symbols for the mass fractions. There does not seem to be an agreed-upon standard.

Solution: Using eqn. (10.26), we have

$$\Delta v = -g_c I_{\rm sp} \ln[\beta + (1-\beta)\pi]$$

which implies that the velocity is maximized when $\pi = 0$. Then the velocity is simply

$$\Delta v = -g_c I_{\rm sp} \ln[\beta] - 230 g_c \ln[0.08] = 581 g_c$$

which comes out to 18690 feet per second or 5697 meters per second. ♠

Note that this example implies that such a rocket could never be used to attain orbit from the Earth's surface, since a low-Earth orbit requires about 7.7 km/s of orbital speed (and about another 1.5 km/s that gets lost in overcoming drag and gravity). This leads to the multi-stage rocket, the subject of the next section.

10.6 Multi-Stage Rockets

The difficulty with using single-stage rockets to achieve space is the amount of structural mass that must be carried along. To alleviate this problem, it is advisable to discard excess mass when possible. This is accomplished through *staging*, in which several rocket motors are used in turn, each being discarded when its fuel is exhausted. In this way, the structural mass of the expended motor is also done away with, and the necessity to further accelerate it eliminated.

In this section, we will consider only the most basic form of staging, in which only one motor is fired at a time, and is discarded before the next is ignited. More general forms of staging are often used. The space shuttle, for example, fires both the main engine and the strap-on boosters during launch. The boosters are dropped when their fuel is expended, but the main engine remains. The external fuel tank is later dropped, as well. Thus the shuttle has in a sense three stages (since it has different structural mass in each phase), but one engine is active in all three.

10.6.1 Two-Stage Rockets

Consider a rocket with only two stages. The stages are stacked one atop the other, and numbered in the order in which they are fired, so that stage one is the lower stage. In this simplest case, the multi-stage rocket can be considered two single-stage vehicles. The second stage is a usual rocket, with structural mass $m_{s,2}$, propellant $m_{p,2}$, and payload P. The first stage has structural and propellant masses $m_{s,1}$ and $m_{p,1}$, and its payload is everything it carries. Thus the payload of the first stage is the total initial mass of the second stage. The payload mass fractions of the two stages are then

$$\pi_1 = \frac{m_{0,2}}{m_{0,1}}; \qquad \pi_2 = \frac{P}{m_{0,2}}$$

The overall payload fraction of the vehicle is the payload over the initial mass of the vehicle. The initial mass of the vehicle is the initial mass of the first stage (since its initial mass includes its payload, which in turn includes all other stages), so

$$\pi^{\star} = \frac{P}{m_{0,1}} = \frac{m_{0,2}}{m_{0,1}} \frac{P}{m_{0,2}} = \pi_1 \pi_2 \tag{10.27}$$

We can use the same equations to find the performances of the two stages that we had for the single-stage rocket. Given the payload fractions and structural factors of the two stages, we use eqn. (10.26) for each stage to get

$$\begin{aligned} \Delta v_1 &= -g_c I_{sp_1} \ln[\beta_1 + (1-\beta_1)\pi_1] \\ \Delta v_2 &= -g_c I_{sp_2} \ln[\beta_2 + (1-\beta_2)\pi_2] \end{aligned}$$

and the total velocity achieved is simply

$$\Delta v = \Delta v_1 + \Delta v_2$$

In general, there is no reason to expect that the mass fractions of the two stages will be the same. Quite often, the lowest stage of a vehicle will be a solid rocket, with the topmost stage having a liquid motor.

Example 10.4. We wish to accelerate a payload P to a velocity of 9000 m/s. We will use a booster technology which has a structural factor $\beta = 0.08$, and fuel with specific impulse $I_{\text{sp}} = 230$ seconds. Using a two-stage vehicle, how much propellant will be needed (in terms of P)?

Solution: Since the specific impulse and structural factors are the same, the optimal solution will have the payload fractions and Δv's the same for the two stages. Therefore

$$\begin{aligned} &\Delta v_2 = 4500 = -9.80665 \cdot 230 \ln[0.08 - 0.92\pi] \\ &\implies \quad 0.08 + 0.92\pi = e^{-1.9951} = 0.136 \quad \implies \quad \pi = 0.0609 \end{aligned}$$

For the upper stage, we have

$$m_{0,2} = P/\pi \quad \implies \quad m_{0,2} = 16.43P$$

Now, we know that

$$m_s = \frac{\beta}{1-\beta} m_p \quad \implies \quad m_s = 0.0870 m_p$$

so we have that

$$\begin{aligned} &P + m_p + 0.0870 m_p = 16.43P \\ &\implies \quad 1.0870 m_p = 15.43P \quad \implies \quad m_p = 14.2P \end{aligned} \tag{10.28}$$

For the lower stage, the analysis is identical, except that the payload is the entire upper stage. Thus we substitute P/π for P in eqn. (10.28) to get

$$m_{p,1} = 14.2(P/\pi) = 233.2P$$

Finally, we sum the propellant in the two stages to get a total of

$$m_p = 14.2P + 233.2P = 247.4P$$

♠

It may seem that almost 250 kg of propellant for every kg of payload is quite a lot. We saw in example 10.3, however, that the job is not possible at all for a single-stage vehicle.

10.6.2 The n-Stage Rocket

A rocket with any given number n of stages is treated by extending the results for a two-stage rocket. The total velocity of the rocket is the sum of the stage Δv's:

$$\Delta V = -g_c \sum_{i=1}^{n} I_{sp_i} \ln[\beta_i + (1-\beta_i)\pi_i] \tag{10.29}$$

and the overall mass fraction is the product of the stage mass fractions:

$$\pi^\star = \prod_{i=1}^{n} \pi_i \tag{10.30}$$

Again, there is no reason that the stages of the rocket should be identical, particularly given the differing technologies that may be used for them. In the very special case in which all of the stages use the same technology, so that the structural factors and specific impulses are the same, it turns out that the most efficient way to construct the rocket is to have the same payload fraction for each stage, and thus the same Δv for each stage. By "most efficient", we mean that the rocket will achieve the greatest velocity for a given final payload and specified amount of fuel.

Example 10.5. For a specified payload P, compute the amount of fuel required to achieve a velocity of 8 km/sec for

1. a two-stage rocket with $I_{sp} = 210$ for each stage
2. a three-stage rocket with $I_{sp} = 210$ for each stage

Assume that π and Δv are the same for each stage, and use $\beta = 0.08$ for all stages.

: Solution Since Δv is the same for all stages, we can write for (a)

$$\Delta v_1 = \Delta v_2 = 4000 = -I_{sp} g_c \ln[\beta + (1-\beta)\pi]$$
$$\implies \quad e^{-1.9423} = 0.08 + 0.92\pi \quad \implies \quad \pi = 0.06888$$

Now, we know

$$\beta = \frac{m_s}{m_s + m_p} \quad \Longrightarrow \quad m_s = \frac{\beta}{1-\beta} m_p$$

Since $\pi = \pi_k = P_k / m_{0,k}$ for the k-th stage, we can use this to get

$$\begin{aligned} m_0 &= m_s + m_p + P = P/\pi \\ &\Longrightarrow \quad (1 + \beta/(1-\beta)) m_{p,k} = P_k/\pi_k \end{aligned}$$

which in this case gives

$$m_{p,k} = 12.436 P_k$$

For the first stage, $P_1 = m_{0,2} = P/\pi = 14.518P$, so

$$m_p = m_{p,1} + m_{p,2} = 12.436 P_1 + 12.436 P_2 = 12.436(1/\pi + 1)P = 193P$$

For part (b), everything is the same, except that we have three stages. Then

$$\Delta v = 8000/3 = 2667 \frac{\text{km}}{\text{s}} \quad \Longrightarrow \quad \pi = 0.2108$$

And

$$(1/\pi - 1)P = (1 + \beta/(1-\beta)) m_p \quad \Longrightarrow \quad m_{p,k} = 3.444 P_k$$

Since we have three stages, we have three terms in the total mass:

$$m_p = 3.444(1/\pi^2 + 1/\pi + 1)P = 97.3P$$

Note that the amount of propellant required per unit mass of payload has been greatly reduced in the 3-stage vehicle. ♠

From example 10.5, we can see that increasing the number of stages greatly increases the efficiency of a rocket. Adding a fourth stage would reduce the propellant requirement still further. The limiting case would be a rocket made up of an infinite number of infinitely small boosters. While this case is clearly infeasible, it is interesting to consider.

The defining characteristic of such a rocket is that the excess mass is discarded effectively continuously. Thus, we can think of the structure as being "burned" along with the fuel.[6] As such, we can examine the performance by adding the mass of the structure to that of the expelled fuel, without increasing the thrust.

Recalling that

$$m_s = \frac{\beta}{1-\beta} m_p$$

we can define the mass of propellant and structure, calling it $\tilde{m}_p$, as

$$\tilde{m}_p = m_s + m_p = \frac{1}{1-\beta} m_p$$

[6] Such designs have in fact been proposed; so far as the author is aware, none have been developed beyond preliminary examination.

We also have that $T = u\dot{m}_p$, and from the definition of specific impulse that $u = I_{sp}g_c$. Therefore

$$T = g_c I_{sp} \dot{m}_p$$

Since the thrust will be unchanged even though the mass flow rate is increased to include the structure, we can adjust the specific impulse to get an effective value $\tilde{I}_{sp}$:

$$T = g_c I_{sp} \dot{m}_p = g_c \tilde{I}_{sp} \dot{\tilde{m}}_p = g_c \tilde{I}_{sp} \frac{\dot{m}_p}{1-\beta} \quad \Longrightarrow \quad \tilde{I}_{sp} = (1-\beta) I_{sp}$$

We can then use the rocket equation, with $\beta = 0$, to derive a maximum possible value for Δv. Carrying out the arithmetic, we get

$$\Delta v = -g_c \tilde{I}_{sp} \ln[\pi]$$

If we also took the payload ratio to zero, we would get infinite velocity; this is not only ludicrous, but useless.

10.7 Problems

1. At $t = 0$ the engine of a single-stage rocket ignites. Due to a design flaw, the thrust is only half the initial weight of the rocket. In terms of I_{sp}, when will the rocket actually lift off? (Assume constant mass flow rate.)

2. The total Δv required for transfer from low earth orbit with 28° inclination to geosynchronous equatorial orbit is 4.49 km/s. A space tug with a restartable engine has $I_{\text{sp}} = 453\,\text{s}$, 16,000 kg of fuel, and structural mass of 1300 kg.

   (a) How much payload can the tug deliver to geosynchronous orbit?

   (b) Can the tug make the round trip without payload? If so, how much payload can it carry to geosynchronous orbit and still make it back (after leaving the payload in orbit) to low earth orbit?

3. Consider the case of a vehicle with two strap-on boosters, such as the Space Shuttle. The solid fuel gives the boosters a specific impulse of $I_{\text{sp}} = 287\,\text{s}$, and the booster technology is such that the structural factor $\beta = 0.148$. The main vehicle has a liquid fuel engine with $I_{\text{sp}} = 455\,\text{s}$ that generates 1.41 million pounds of thrust.

   (a) If the burn time for the boosters is $t_b = 120\,\text{s}$ and each produces 2.65 million pounds of thrust, what is the weight of the boosters, both initially (before ignition), and after the fuel is expended?

(b) When the vehicle is used, the boosters fire along with the main engine. When the boosters burn out, their empty shells are dropped while the main engine continues to fire. If the main engines fire for a total of eight minutes, and the vehicle weighs 375,000 lb. when they shut off, what is the total Δv capability of the vehicle?

4. A two-stage sounding rocket in which both stages have $I_{\rm sp} = 282\,\rm s$ carries 1167 kg of propellant in the first stage and 415 kg in the second. The structural masses of the two stages are 113 kg and 41 kg, and the vehicle carries a payload of 250 kg.

   Ignoring the effects of burn time (see note below),

   (a) Calculate the approximate burnout speed, and the final altitude, of the vehicle.

   (b) Instead of the usual strategy, assume instead that we allow the second stage to coast after the first stage burns out, until it reaches its maximum altitude, and then we ignite the second stage. What is the resulting final altitude?

   Note: To ignore the effects of burn time, simply ignore the gt term in the sounding rocket boost equation. This is equivalent to assuming that the altitude gained during burn is negligible compared to the altitude gained during coast.

Chapter 11

Solar Sailing

So far, this is just a filler.

Bibliography

[1] Bate, R.R., Mueller, D.D., and White, J. E., *Fundamentals of Astrodynamics*, Dover Publications, New York, 1971.

[2] Battin, Richard H., *An Introduction to the Mathematics and Methods of Astrodynamics*, Revised edition, AIAA Education Series (J. S. Przemieniecki, Series Ed.), AIAA, Inc, Reston, Virginia, 1999.

[3] Berry, Arthur, *A Short History of Astronomy: From Earliest Times through the Nineteenth Century*, Dover Publications, New York, 1961 (republication of work orginally published by John Murray, 1898).

[4] Brouwer, Dirk, and Gerald M. Clemence, *Methods of Celestial Mechanics*, Academic Press, New York, NY, 1961.

[5] Curtis, Howard D., *Orbital Mechanics for Engineering Students*, Elsevier Butterworth_Heinemann, Jordan Hill, Oxford, 2005.

[6] Danby, J.M.A., *Fundamentals of Celestial Mechanics*, 2nd ed., Willman-Bell, Inc, Richmond, Virginia, 1992.

[7] Gurzadyan, G.A., *Theory of Interplanetary Flights*, Gordon and Breach Publishers, The Netherlands, 1996.

[8] Kaplan, Marshall H., *Modern Spacecraft Dynamics and Control*, John Wiley and Sons, New York, 1976.

[9] Kellogg, Oliver Dimon, *Foundations of Potential Theory*, Dover Publications Inc., 1954.

[10] Meeus, Jean, *Astronomical Algorithms*, 2$^{\text{nd}}$ edition, Willman-Bell, Richmond, Virginia, 1998.

[11] Moulton, F.R., *An Introduction to Celestial Mechanics,* 2nd revised ed., Dover, New York, 1970 (republication of work originally published by the MacMillan Company, 1914).

[12] Prussing, J.E., and Conway, B. A., *Orbital Mechanics,* Oxford, New York, 1993.

[13] Roy, A.E., *Orbital Motion*, 3rd ed., Adam Hilger, Bristol, 1988.

[14] Sutton, George P., *Rocket Propulsion Elements: An Introduction to the Engineering of Rockets,* 5th edition, John Wiley and Sons, New York, 1986.

[15] Thomson, William Tyrrell, *Introduction to Space Dynamics*, Dover Publications, Inc., New York, 1986.

[16] Vallado, David A., *Fundamentals of Astrodynamics and Applications*, 2nd edition, Space Technology Library Volume 12, Microcosm Press, El Segindo, California, 2001.

[17] Wiesel, William E., *Spaceflight Dynamics,* McGraw-Hill, Inc., New York, 1989.

Appendix A

Sidereal Time and Various Years

There is nothing in orbital mechanics, at least not to the extent that it is covered in this book, that makes the passage of time confusing. However, the *telling* of time is not always straightforward when it comes to defining days and years.

A.1 Solar vs. Sidereal Days

The first difficulty in defining time comes from defining exactly what is a day. Classically, this was not a problem, as the Earth was thought to be the center of the universe and the Sun revolved about it. A day was then the length of time between noons; that is, from when the Sun was highest in the sky until it was next highest in the sky.

This was simple enough and more than sufficient until it became known that the Earth in fact moved around the Sun.[1] This leads to the use of the fixed stars as the frame of reference, as in Chapter 4. In this frame, the day is more reasonable defined as the time between the highest points of some fixed star or other celestial object. In fact, the *sidereal day* is defined as the time between successive passages of the first point of Aries over the observer's longitude. This is a slightly shorter length of time than the *solar day* discussed above.

The term *sidereal* means "with respect to the stars". The fact is that the Earth does not rotate about 365 and one-fourth times per year. Instead, it rotates once more, when viewed with respect to the fixed stars. This is because during each day the Earth also moves about the Sun by about one degree, and so the Earth must rotate that much further to bring a point on its surface once again between the center

[1]It is of course equally true to say that the Sun revolves around the Earth, as it is now known that there can be no true center defined to the universe. However, the other planets do not revolve around the Earth, and it is certainly more *convenient* to view the Sun as the center.

of the Earth and the Sun. As the Earth does not move detectably with respect to the distant stars, there is no corresponding increase in the rotation needed to bring one of the stars back to an overhead position. Therefore the Earth rotates 360 degrees for each sidereal day, and approximately 360 + 360/365.25 degrees in the average solar day.

A.1.1 The Mean Solar Day

Because the Earth moves in an elliptic orbit, it does not move through the same angle every day. The solar day thus has a slightly different length at different times of the year. While the variation is small, it is measurable. This leads to the definition of the *mean solar day*, which is the average length of a day during one sidereal year.[2]

This is more important than it may at first seem, because the second is defined as 1/86400 of a mean solar day.[3] This is the usual second. There is also a *sidereal second*, which is analogous to the mean solar second; it is 1/86400 of a sidereal day.

A.2 Tropical and Sidereal Years

The year as usually understood is the *sidereal year*, which is to say when the Sun has regained its position with respect to the fixed stars (as seen from the Earth). Recall that the heliocentric frame is defined in terms of the vernal equinox direction (sec 4.1), and that this direction is not quite constant. Therefore, the sidereal year is not quite the same as the *tropical year*, which is the time between successive passages of the Sun across the vernal equinox. Still another year is the *anomalistic year*, which is the time from perihelion to perihelion of the Earth's orbit; due to various third-body effects, this is not quite constant.

---

[2] This variation is on the order of seconds. The Earth moves just under 1.02° during a mean solar day at perihelion, versus a bit over 0.953° at aphelion, and it takes about 16 seconds for the Earth to rotate through the additional angle.

[3] Actually, it is now defined as some number of wavelengths of some radiation of something, which can be precisely measured in an atomic clock. But the nember was chosen to come out to 1/86400 of a mean solar day, as usually understood.

Appendix B

Review of Vectors and Kinematics

In this section we present a basic review of vector arithmetic and some results from kinematics. This is not intended to be a complete treatment of the subject. For a more complete review, see Thomson [15], *Introduction to Space Dynamics*, chapters 1 through 3.

B.1 Vector Arithmetic

For our purposes, a vector will usually denote a quantity in 3 dimensions, such as a position or velocity in Cartesian coordinates. Thus the review given here will be confined to finite-dimensional vectors in general, and in some cases specifically to 3-dimensional vectors. Except when specifically noted, the same definitions and operations apply to all finite-dimensional vectors.

B.1.1 Notation

We will consider column vectors, and our examples will be 3-dimensional, such as

$$\mathbf{a} = \left\{ \begin{array}{c} a_1 \\ a_2 \\ a_3 \end{array} \right\} \qquad \mathbf{b} = \left\{ \begin{array}{c} b_1 \\ b_2 \\ b_3 \end{array} \right\}$$

We will in general use the bold face lower-case Roman character to denote a vector, and the subscripted character to denote a member of the vector. The plain character without subscript will be used to denote the magnitude of the vector (the definition of magnitude will be given later).

The *zero vector* is the vector consisting of all zeroes. It is usually written $\mathbf{0}$, but it is not uncommon to see notation such as $\mathbf{a} = 0$, as well.

B.1.2 Addition and Scalar Multiplication

The addition of vectors is defined as the addition of the entries in the vectors. Thus we have

$$\mathbf{c} = \mathbf{a} + \mathbf{b} = \left\{ \begin{array}{c} a_1 + b_1 \\ a_2 + b_2 \\ a_3 + b_3 \end{array} \right\} = \left\{ \begin{array}{c} c_1 \\ c_2 \\ c_3 \end{array} \right\}$$

The multiplication of a vector by a scalar is also defined entry-by-entry as follows:

$$k\mathbf{a} = \left\{ \begin{array}{c} ka_1 \\ ka_2 \\ ka_3 \end{array} \right\}$$

Note that each of these operations is commutative; that is, $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$ and $k\mathbf{a} = \mathbf{a}k$.

B.1.3 Magnitude

The *magnitude* (or *norm*) of a vector means, in simple terms, its length. For our purposes, the magnitude is the Euclidean norm, defined for a vector of arbitrary finite dimension as the square root of the sum of the squares of the entries in the vector. Using a to denote the magnitude of the vector $\mathbf{a}$, we thus have

$$a = \sqrt{\sum_{i=1}^{n} a_i^2}$$

where n is the number of entries in the vector. For a vector with three entries, this is often seen written as

$$a = \left(a_1^2 + a_2^2 + a_3^2\right)^{\frac{1}{2}}$$

A common notation for the norm of a vector is also $\|\mathbf{a}\|$.

It should be remembered that the definition we will use is not the only correct definition of norm. In a more general mathematical sense, a norm need only satisfy some given requirements. This is both beyond the scope of this review and of little immediate use.

B.1.4 Expression through Unit Vectors

A *unit vector* is simply a vector with length 1. A unit vector can be derived from any vector $\mathbf{a}$ (except the zero vector), simply by dividing the vector by its magnitude:

$$\hat{\mathbf{a}} = \mathbf{a}/a$$

The result is a vector of unit length, pointing in the same direction as the original vector. It is immediately obvious that $\mathbf{a} = a\hat{\mathbf{a}}$. (The use of the circumflex ('hat') to denote a unit vector is fairly standard.)

From the last section, it is clear that we can write any three-dimensional vector $\mathbf{a}$ as

$$\mathbf{a} = \left\{ \begin{array}{c} a_1 \\ a_2 \\ a_3 \end{array} \right\} = \left\{ \begin{array}{c} a_1 \\ 0 \\ 0 \end{array} \right\} + \left\{ \begin{array}{c} 0 \\ a_2 \\ 0 \end{array} \right\} + \left\{ \begin{array}{c} 0 \\ 0 \\ a_3 \end{array} \right\} = a_1 \left\{ \begin{array}{c} 1 \\ 0 \\ 0 \end{array} \right\} + a_2 \left\{ \begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right\} + a_3 \left\{ \begin{array}{c} 0 \\ 0 \\ 1 \end{array} \right\}$$

The expression to the right of the second equals sign expresses $\mathbf{a}$ as the sum of the three component values a_1, a_2, and a_3, each multiplied by a unit vector that causes the component to take its proper place in $\mathbf{a}$. The set of unit vectors shown is usually denoted $\hat{\imath}, \hat{\jmath}$, and $\hat{k}$.

When the vector is expressed in a rectangular coordinate frame, these three unit vectors point along the three coordinate axes.

B.1.5 Dot and Cross Products

The *dot product* (sometimes referred to as the *scalar product*) of two vectors is defined as

$$\mathbf{a} \cdot \mathbf{b} = ab \cos \phi$$

where ϕ is the angle between the two vectors. It is immediate that

$$\mathbf{a} \cdot \mathbf{a} = aa \cos 0 = a^2$$

Because the unit vectors $\hat{\imath}, \hat{\jmath}$, and $\hat{k}$ are aligned with the coordinate axes, the angles between them are 90°, and we have that

$$\hat{\imath} \cdot \hat{\jmath} = \hat{\imath} \cdot \hat{k} = \hat{\jmath} \cdot \hat{k} = 0$$

From previous sections, we can write

$$\mathbf{a} \cdot \mathbf{b} = \left(a_1\hat{\imath} + a_2\hat{\jmath} + a_3\hat{k} \right) \cdot \left(b_1\hat{\imath} + b_2\hat{\jmath} + b_3\hat{k} \right)$$

which multiplies out to give

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} \quad = \quad & a_1 \left(b_1\hat{\imath} \cdot \hat{\imath} + b_2\hat{\imath} \cdot \hat{\jmath} + b_3\hat{\imath} \cdot \hat{k} \right) + a_2 \left(b_1\hat{\jmath} \cdot \hat{\imath} + b_2\hat{\jmath} \cdot \hat{\jmath} + b_3\hat{\jmath} \cdot \hat{k} \right) \\ & + a_3 \left(b_1\hat{k} \cdot \hat{\imath} + b_2\hat{k} \cdot \hat{\jmath} + b_3\hat{k} \cdot \hat{k} \right) \quad = \quad a_1b_1 + a_2b_2 + a_3b_3 \end{aligned}$$

More generally, for vectors of length n, we write

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{n} a_ib_i$$

From this, it is immediately obvious that the dot product is commutative, and distributive over addition. That is,

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a} \qquad \text{and} \qquad (\mathbf{a} + \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} + \mathbf{b} \cdot \mathbf{c}$$

The *cross* (or *vector*) *product* is defined only for vectors of length 3. It is defined as

$$\mathbf{a} \times \mathbf{b} = (ab \sin \phi)\hat{\mathbf{c}}$$

where $\hat{\mathbf{c}}$ is a unit vector perpendicular to both $\mathbf{a}$ and $\mathbf{b}$ and positive by the right-hand rule. The right-hand rule is defined according to the order of the vectors: the direction of $\hat{\mathbf{c}}$ is obtained by first aligning the fingers of the right hand with $\mathbf{a}$, holding the hand so that the fingers can curl into the direction of $\mathbf{b}$. $\hat{\mathbf{c}}$ is then in the direction of the thumb.

Since the cross product gives $\hat{\mathbf{c}}$ in opposite directions for $\mathbf{a} \times \mathbf{b}$ and $\mathbf{b} \times \mathbf{a}$, we see that the cross product is not commutative, and in fact that $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$. The cross product is, however, distributive over addition.

It is obvious from the definition that $\mathbf{a} \times \mathbf{a} = 0$.

B.2 Coordinate Transformations

In this section, we will consider only three-dimensional coordinate frames, and will assume that the origin of each frame is at the same point. We are interested in the ways that a single vector may be expressed in each frame, and how the two expressions are related.

B.2.1 Single-axis Rotations

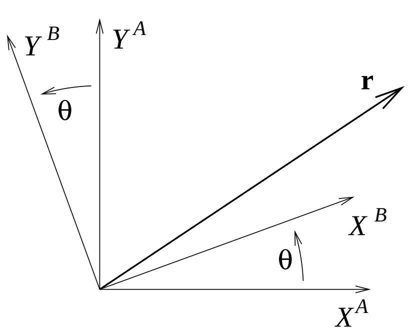


Figure B.1: Rotation about the Z-axis.

Consider the situation of Figure B.1. The sketch shows two coordinate frames, with a common origin, both with the Z-axis out of the page. The two are related by the rotation angle ϕ; the frame superscripted B can be obtained by rotating the frame superscripted A through ϕ about the Z-axis.

Suppose that the vector $\mathbf{r}$ is expressed in the original frame as

$$\mathbf{r} = r_1^A \hat{\imath}^A + r_2^A \hat{\jmath}^A + r_3^A \hat{k}^A.$$

The same vector will be expressed in the rotated frame as

$$\mathbf{r} = r_1^B \hat{\imath}^B + r_2^B \hat{\jmath}^B + r_3^B \hat{k}^B$$

where of course the new set of unit vectors are aligned with the rotated axes. Now, by definition $r_1^B = \mathbf{r} \cdot \hat{\imath}^B$. We can thus write

$$r_1^B = \left(r_1^A \hat{\imath}^A + r_2^A \hat{\jmath}^A + r_3^A \hat{k}^A\right) \cdot \hat{\imath}^B = r_1^A \hat{\imath}^A \cdot \hat{\imath}^B + r_2^A \hat{\jmath}^A \cdot \hat{\imath}^B + r_3^A \hat{k}^A \cdot \hat{\imath}^B \tag{B.1}$$

and recalling the definition of the dot product, we can express this as

$$\begin{aligned} r_1^B &= r_1^A \cos\phi + r_2^A \cos(\pi/2 - \phi) + r_3^A \cos(\pi/2) \\ &= r_1^A \cos\phi + r_2^A \sin\phi + r_3^A \end{aligned}$$

Similarly, we have

$$\begin{aligned} r_2^B &= \left(r_1^A \hat{\imath}^A + r_2^A \hat{\jmath}^A + r_3^A \hat{k}^A\right) \cdot \hat{\jmath}^B \\ &= r_1^A \hat{\imath}^A \cdot \hat{\jmath}^B + r_2^A \hat{\jmath}^A \cdot \hat{\jmath}^B + r_3^A \hat{k}^A \cdot \hat{\jmath}^B \\ &= r_1^A \cos(\pi/2 + \phi) + r_2^A \cos\phi + r_3^A \\ &= -r_1^A \sin\phi + r_2^A \cos\phi + r_3^A \end{aligned}$$

And obviously $r_3^B = r_3^A$. Putting these into matrix form, we have

$$\left\{ \begin{array}{c} r_1^B \\ r_2^B \\ r_3^B \end{array} \right\} = \left[\begin{array}{ccc} \hat{\imath}^A \cdot \hat{\imath}^B & \hat{\jmath}^A \cdot \hat{\imath}^B & \hat{k}^A \cdot \hat{\imath}^B \\ \hat{\imath}^A \cdot \hat{\jmath}^B & \hat{\jmath}^A \cdot \hat{\jmath}^B & \hat{k}^A \cdot \hat{\jmath}^B \\ \hat{\imath}^A \cdot \hat{k}^B & \hat{\jmath}^A \cdot \hat{k}^B & \hat{k}^A \cdot \hat{k}^B \end{array} \right] \left\{ \begin{array}{c} r_1^A \\ r_2^A \\ r_3^A \end{array} \right\} \tag{B.2}$$

$$= \left[\begin{array}{rrr} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{array} \right] \left\{ \begin{array}{c} r_1^A \\ r_2^A \\ r_3^A \end{array} \right\} \tag{B.3}$$

Using the notation $\mathbf{r}^A$ to mean the vector $\mathbf{r}$ expressed in relation to the unit vectors of frame A, we can write

$$\mathbf{r}^B = L_{BA} \mathbf{r}^A$$

where the *rotation matrix* L_{BA} transforms the coefficients of $\hat{\imath}^A, \hat{\jmath}^A$, and $\hat{k}^A$ to those multiplying $\hat{\imath}^B, \hat{\jmath}^B, \hat{k}^B$. Note that despite the different way of expressing it, the matrix clearly must be the same in both (B.2) and (B.3).

Now, a vector described in the rotated frame B can be related to the unit vectors of frame A in an identical fashion. To use the form of (B.2), we have

$$\left\{ \begin{array}{c} r_1^A \\ r_2^A \\ r_3^A \end{array} \right\} = \left[\begin{array}{ccc} \hat{\imath}^B \cdot \hat{\imath}^A & \hat{\jmath}^B \cdot \hat{\imath}^A & \hat{k}^B \cdot \hat{\imath}^A \\ \hat{\imath}^B \cdot \hat{\jmath}^A & \hat{\jmath}^B \cdot \hat{\jmath}^A & \hat{k}^B \cdot \hat{\jmath}^A \\ \hat{\imath}^B \cdot \hat{k}^A & \hat{\jmath}^B \cdot \hat{k}^A & \hat{k}^B \cdot \hat{k}^A \end{array} \right] \left\{ \begin{array}{c} r_1^B \\ r_2^B \\ r_3^B \end{array} \right\} \tag{B.4}$$

It is obvious that the matrix in (B.4) is the transpose of that in (B.2). We could come to the same conclusion by noting that the tranformation from B to A would require a rotation through $-\phi$ about Z. Substituting $-\phi$ for ϕ in (B.3) produces

$$\left\{ \begin{array}{c} r_1^A \\ r_2^A \\ r_3^A \end{array} \right\} = \left[\begin{array}{rrr} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{array} \right] \left\{ \begin{array}{c} r_1^B \\ r_2^B \\ r_3^B \end{array} \right\}$$

which again is the transpose of the earlier matrix.

B.2.2 Euler Rotations

Consider any two rectilinear coordinate systems with a common origin. Euler showed that the axes of one of them can be rotated to align with those of the other by not more than three rotations, each about a unique axis. This result is usually implemented by defining the three axis in advance. The values of the three rotation angles then fully describe the orientation of one coordinate system with respect to the other. The angles are often referred to as the *Euler angles*.

B.3 Kinematics

We now review some basic facts and defintions from kinematics, the study of objects in motion. We will generally consider motion in three dimensions, but will also often restrict the motion to plane. Planar motion can be considered a special case of three-dimensional motion, and all general statements about three-dimensional motion apply to the simpler case.

B.3.1 Position and Velocity

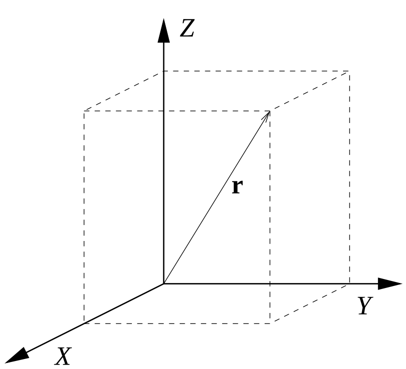


Figure B.2: Position vector.

For our purposes, we will consider motion in a rectangular coordinate frame, and we will usually denote the axes as X, Y, and Z. The position of some point in the frame is described by a vector of length three, as

$$\mathbf{r} = \left\{ \begin{array}{c} r_1 \\ r_2 \\ r_3 \end{array} \right\} = \left\{ \begin{array}{c} x \\ y \\ z \end{array} \right\} = x\hat{\imath} + y\hat{\jmath} + z\hat{k}$$

where the three entries are the displacement of the point from the origin along the X, Y, and Z axes, respectively.[1]

[1]The notation for position and velocity is variable. It is common to see r_x to mean the x-component of $\mathbf{r}$, and r_i for the same quantity (the 'i' from the unit vector $\hat{\imath}$) is equally common.

The *velocity* of the point in the frame is by definition the time rate of change of the position:

$$\mathbf{v} \triangleq \frac{d\mathbf{r}}{dt} = \dot{\mathbf{r}}$$

Recalling the usual definition of a derivative, we can write this as

$$\mathbf{v} = \lim_{\Delta t \to 0} \frac{\mathbf{r} + \Delta\mathbf{r}}{\Delta t}$$

where Δt is a scalar. Since division of a vector by a scalar is done element-by-element, it follows that derivatives are taken element-by-element, as well.

Note that while $\mathbf{v} = \dot{\mathbf{r}}$, it is *not* true that $v = \dot{r}$.

The acceleration is defined as the time rate of change of velocity.

B.3.2 Motion in a Rotating Frame

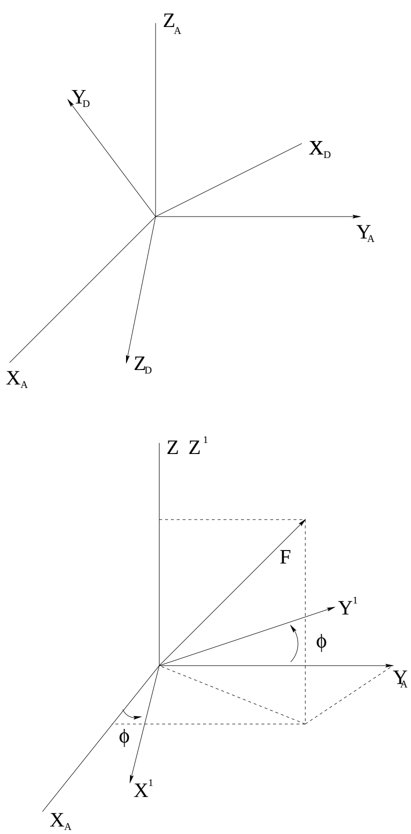


Figure B.3: PLEASE FIND A NAME FOR ME

In our kinematics review, we occasionally expressed the same vector in different coordinate systems. This is what a coordinate transformation does. We will explore the topic in general now.

Suppose we have two coordinate systems, say A and D, and we will assume that they are coincident - the origins are at the same point.

It can be shown that the axes of D can be rotated into alignment with A and vice versa, through a series of 3 angular rotations. There are many possible rotations we could use. Also, even using the same rotations, the order is important; different orders produce different results.

Suppose that we first rotate about the vertical axis through an angle ϕ. Alternatively, consider the frame $X_B Y_B Z - B$, in which Z_B is collinear with Z. Let $\mathbf{r}$ be some vector, which we write as

$$\mathbf{r} = r_1 \hat{i}_A + r_2 \hat{j}_A + r_3 \hat{k}_A$$

How would $\mathbf{r}$ be expressed in the he $X^1 Y^1 Z^1$ system? We know that we can write

$$\mathbf{r} = (\mathbf{r} \cdot \hat{i}^1)\hat{i}^1 + (\mathbf{r} \cdot \hat{j}^1)\hat{j}^1 + (\mathbf{r} \cdot \hat{k}^1)\hat{k}^1$$

Consider the first of these. Expanding we have

$$\mathbf{r} = (r_1\hat{i}_A \cdot \hat{i}^1)\hat{i}^1 + (r_2\hat{j}_A \cdot \hat{i}^1)\hat{i}^1 + (r_3\hat{k}_A \cdot \hat{i}^1)\hat{i}^1 + \ldots$$

Now, $\hat{i}_A \cdot \hat{i}^1 = |\hat{i}_A| \cdot |\hat{i}^1| \cos\phi$. Similarly, $\hat{j}_A \cdot \hat{i}^1 = +\sin\phi$. And, since $\hat{k}_A$ is normal to $\hat{i}^1$, we get $\hat{k}_A \cdot \hat{i}^1 = 0$.

Expanding for $\hat{j}^1$ and $\hat{k}^1$ terms gives

$$\begin{aligned} \mathbf{r} &= r_1 \cos\phi \hat{I}^1 + r_2 \sin\phi \hat{i}^1 \\ &- r_2 \sin\phi \hat{j}^1 + r_2 \cos\phi \hat{j}^1 + r_3 \hat{k}^1 \end{aligned}$$

and this is, of course, the expression of $\mathbf{r}$ in the B coordinate system. We can write this in vector-matrix form as

$$\mathbf{B}_r = \begin{Bmatrix} r_{1B} \\ r_{2B} \\ r_{3B} \end{Bmatrix} = \begin{bmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} r_{1A} \\ r_{2A} \\ r_{3A} \end{Bmatrix} = L_{BA}\mathbf{r}_A$$

(The subscript on L means "A to B".) L_{AB} is the transformation matrix from A to B.

Suppose that instead of rotating about Z, we had rotated through an angle θ about Y. Our sketch might look like

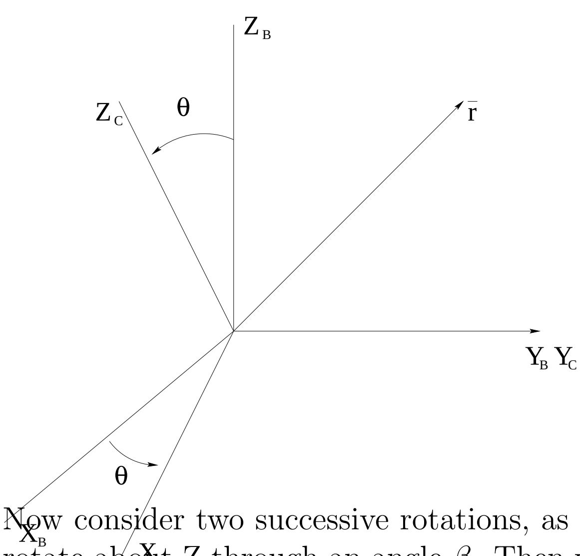


Figure B.4: I AM NOTHING WITH OUT A NAME

and through manipulation similar to those we just performed, we find that

$$\mathbf{r}_c = \begin{bmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{bmatrix} \mathbf{r}_B \Longrightarrow \mathbf{r}_C = L_{CB}\mathbf{r}_B$$

A rotation about an X axis, say through an angle ψ, would give the transformation

$$\mathbf{r}_D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{bmatrix} \mathbf{r}_C = L_{DC}\mathbf{r}_C$$

Now consider two successive rotations, as with a radar dish or an aircraft. First, we rotate about Z through an angle β. Then we rotate about the moved Y axis through the angle α. (In actuality of course, we would do the rotations simultaneously to follow a target. But we can do them separately for the purpose of initial aiming.)

Now we have a target at some point $\mathbf{r}_C$ in the radar frame. Looking at the process of moving from one frame (the inertial) to another, we can define an intermediate frame,B, after the first rotation. It is in this intermediate frame that the second rotation is performed.

Since we have an intermediate frame, we can express **r** in that frame. Thus we can write

$$\mathbf{r}_C = L_{CB}\mathbf{r}_B$$

But we also have

$$\mathbf{r}_B = L_{BI}\mathbf{r}_I$$

so

$$\begin{aligned} \mathbf{r}_c &= L_{CB}L_{BI}\mathbf{r}_I \\ \Longrightarrow \mathbf{r}_I &= (L_{CB}L_{BI}^{-1}\mathbf{r}_C \end{aligned}$$

This continues through a third rotation, if necessary. Thus, we can define the complete transformation matrix from A to D through

$$\begin{aligned} \mathbf{r}_D &= L_{CD}L_{CB}L_{BA}\mathbf{r}_A \\ \Longrightarrow L_{DA} &= L_{DC}L_{CB}L_{BA} \end{aligned}$$

It must be stressed that the rotations need not be about and X,Y, and Z axis. We could just as easily use, say ϕ about Z_A, θ about Y_B, and ψ about Z_C. The only rule is that we cannot use the same axis twice, since a rotation about an axis does not move that axis.

B.3.3 Direction Cosines

We noted before that any vector can be written as

$$\mathbf{r}_B = (\mathbf{r}\cdot\hat{i}_B)\hat{i}_B + (\mathbf{r}\cdot\hat{j}_B)\hat{j}_B + (\mathbf{r}\cdot\hat{k}_B)\hat{k}_B$$

and, in particular, that this can be expanded into the form

$$\mathbf{r}_B = \begin{bmatrix} \hat{i}_B\cdot\hat{i}_A & \hat{i}_B\cdot\hat{j}_A & \hat{i}_B\cdot\hat{k}_A \\ \hat{j}_B\cdot\hat{i}_A & \hat{j}_B\cdot\hat{j}_A & \hat{j}_B\cdot\hat{k}_A \\ \hat{k}_B\cdot\hat{i}_A & \hat{k}_B\cdot\hat{j}_A & \hat{k}_B\cdot\hat{k}_A \end{bmatrix}$$

Each entry in this matrix is the cosine of some angle, and these are thus called <u>direction cosines.</u> Note that the transformation from B to A is accomplished by a very similar matrix:

$$\mathbf{r}_A = \begin{bmatrix} \hat{i}_A\cdot\hat{i}_B & \hat{i}_A\cdot\hat{j}_B & \hat{i}_A\cdot\hat{k}_B \\ \hat{j}_A\cdot\hat{i}_B & \hat{j}_A\cdot\hat{j}_B & \hat{j}_A\cdot\hat{k}_B \\ \hat{k}_A\cdot\hat{i}_B & \hat{k}_A\cdot\hat{j}_B & \hat{k}_A\cdot\hat{k}_B \end{bmatrix} = L_{AB}\mathbf{r}_B$$

Examining L_{BA} and L_{AB}, and recalling that $\mathbf{a}\cdot\mathbf{b} = \mathbf{b}\cdot\mathbf{a}$, we see by inspection that

$$L_{BA} = (L_{AB}^T$$

B.3.4 Inversion

Now, we know that

$$\mathbf{r}_A = L_{AB}\mathbf{r}_B \Longrightarrow \mathbf{r}_B = (L_{AB})^{-1}\mathbf{r}_A = L_{BA}\mathbf{r}_A$$

assuming that L_{AB} is invertible, which it always is.

However, we just showed that $L_{AB} = L_{AB}^T$, so we have that

$$(L_{AB})^1 = (L_A B)^T$$

a result that makes life much easier in many cases.

Appendix C

Series Expansions

This chapter presents a brief review of two expansion techniques that will be of use in the text. The first, often known as Taylor's Expansion, is likely to be well-known to the reader. The second technique, attributed to Lagrange, is not is not as well known in general.

C.1 Taylor Expansions

"Taylor Expansions" is the term often used to refer to applications of Taylor's theorem to create a polynomial representation of a function. Begin by reviewing the definition and the simplest properties of the expansion for functions of one variable.

Definition C.1: Consider the function $f(z)$. Let $z = w + \alpha$, where w is fixed. If $f(z)$ has an nth derivative at w its *Taylor expansion of degree n about w* is the polynomial in α written as

$$P(\alpha) = f(w) + f'(w)\alpha + \frac{1}{2!}f''(w)\alpha^2 + \ldots + \frac{1}{n!}f^{(n)}(w)\alpha^n. \tag{C.1}$$

The expansion is valid for any α so long as all the needed derivatives exist on the interval between w and $w + \alpha$. ♠

More formally, Taylor's theorem can be stated as

Theorem C.1: If f has a continuous nth derivative in a neighborhood of w, then in that neighborhood

$$f(z) = f(w) + \frac{1}{1!}f'(w)(z-w) + \ldots + \frac{1}{n!}f^{(n)}(w)(z-w)^n + R_n, \tag{C.2}$$

where

$$R_n = \frac{1}{(n-1)!}\int_w^z (z-t)^{n-1}\left[f^{(n)}(t) - f^{(n)}(w)\right] dt. \tag{C.3}$$

This is often known as the *Integral remainder form* of Taylor's theorem, and other similar names. ∎

The importance of Taylor's theorem lies in the remainder term, which can be shown to go to zero as α goes to zero, and to do so faster than any of the terms of the polynomial. The proof of the theorem is not difficult, but is not necessary for its application, so we omit it.

C.2 Lagrange Expansions

The Lagrange expansion is similar to the Taylor expansion in that it uses the derivatives of the function in the expression. Consider the function $F(z)$ where

$$z = w + \alpha\phi(z). \tag{C.4}$$

Then F can be expressed as a function of w and α as

$$\begin{aligned} F(z) &= F(w) + \alpha\phi(w)F'(w) + \frac{\alpha^2}{2}\frac{\partial}{\partial w}\left[\{\phi(w)\}^2 F'(w)\right] \\ &\quad + \cdots + \frac{\alpha^{n+1}}{(n+1)!}\frac{\partial^n}{\partial w^n}\left[\{\phi(w)\}^{n+1} F'(w)\right] + \cdots \end{aligned} \tag{C.5}$$

Here,

$$F'(z) = \frac{dF}{dz} \quad \Longrightarrow \quad F'(w) = \left.\frac{dF}{dz}\right|_{z=w}.$$

This allows the function to be expressed as an expansion in α about its value at w. It can of course be shown that the Lagrange expansion produces the same expression as does the Taylor expansion, since there can only be one polynomial expansion for any function. It can naturally also be shown that the expansion converges for α sufficiently small.

The Lagrange expansion is particularly useful in astrodynamics because the eccentric anomaly is conveniently expressed as

$$E = M + e\sin E \tag{C.6}$$

so that M takes the place of w, e is the small parameter (α in eqn. C.5) and $\phi(z) = \sin E$. Then

$$\begin{aligned} F(w) &= M \\ F'(\cdot) &= 1 \\ \text{and} \quad \phi(w) &= \sin M. \end{aligned}$$

The expansion for eccentric anomaly about the mean anomaly then becomes

$$E = M + e\sin M + \frac{e^2}{2}\frac{\partial}{\partial M}\left[\sin^2 M\right] + \frac{e^3}{6}\frac{\partial^2}{\partial M^2}\left[\sin^3 M\right] + \cdots \tag{C.7}$$

Taking the derivatives, and appealing to some trigonometric identities, produces

$$\frac{\partial}{\partial M}\left[\sin^2 M\right] = 2\sin M\cos M = \sin 2M \tag{C.8}$$

$$\frac{\partial^2}{\partial M^2}\left[\sin^3 M\right] = 6\sin M\cos^2 M - 3\sin^3 M = \frac{3}{4}\left(3\sin 3M - \sin M\right) \tag{C.9}$$

The first few terms of the expansion are then

$$E = M + e\sin M + \frac{e^2}{2}\sin 2M + \frac{e^3}{8}\left(3\sin 3M - \sin M\right) + \cdots \tag{C.10}$$

The amount of work (largely manipulation of trigonometric identities) increases rapidly with each term in the expansion. However, for small eccentricities the series converges rapidly, and only a few terms are needed.

C.3 Some Trigonometric Identities

Here, we list some trigonometric identities that are used elsewhere in the text. We also give the expansions of some common functions.

C.3.1 Identities

The following are often useful in simplifying expressions involving many trigonometric terms.

$$\sin^2\alpha + \cos^2\alpha = 1$$

$$\begin{aligned}
\sin(\alpha+\beta) &= \sin\alpha\cos\beta + \cos\alpha\sin\beta \\
\sin(\alpha-\beta) &= \sin\alpha\cos\beta - \cos\alpha\sin\beta \\
\cos(\alpha+\beta) &= \cos\alpha\cos\beta - \sin\alpha\sin\beta \\
\cos(\alpha-\beta) &= \cos\alpha\cos\beta + \sin\alpha\sin\beta
\end{aligned}$$

$$\sin^2\alpha = (1-\cos 2\alpha)/2$$
$$\cos^2\alpha = (1+\cos 2\alpha)/2$$

C.3.2 Expansions

In each of the following series, $k = 0, 1, 2, \ldots$.

$$\begin{aligned}
e^x &= 1 + x + x^2/2 + x^3/3! + \cdots + x^k/k! + \cdots \\
\sin x &= x - x^3/3! + x^5/5! + \cdots + (-1)^{k+1}x^{2k+1}/(2k+1)! \\
\cos x &= 1 - x^2/2 + x^4/4! + \cdots + (-1)^k x^{2k}/(2k)!
\end{aligned}$$

Appendix D

A Brief Review of System Stability

This chapter provides a *brief* review of some of the most basic ideas of the stability of dynamical systems. It will in no way serve as a complete introduction to the topic, but should serve to refresh memories and to make comprehensible the ideas that will be presented concerning the stability of the Lagrange points.

We will consider systems of two variables, x and y. These variables will be known as the *state* variables, as they describe the state of the system. Two variables are enough to show the basic ideas of stability, which are applicable to systems of any finite size, while allowing the variables to be dealt with explicitly.

D.1 Problem Formulation and Equilibrium Points

Consider the dynamical system described by the two first-order differential equations

$$\dot{x} = f(x, y) \tag{D.1}$$

$$\dot{y} = g(x, y) \tag{D.2}$$

We will be concerned solely with the stability of *autonomous* systems. The term autonomous means that the independent variable t does not appear explicitly in the equations of motion. The usual spring-mass-damper system is an example of an autonomous system; a rocket, in which the mass changes explicitly with time, is an example of a non-autonomous system.

We will assume that there exists a point $(\tilde{x}, \tilde{y})$ such that the functions f and g above are zero; that is,

$$f(\tilde{x}, \tilde{y}) = g(\tilde{x}, \tilde{y}) = 0$$

This clearly means that the derivatives of the state variables x and y are zero. This is what is meant by the term *equilibrium point*, a point at which the states of the system are constant.

D.2 Motion Near an Equilibrium Point

Supposing we can find a point $(\tilde{x}, \tilde{y})$ as described above, we would like to know how the system behaves in the neighborhood of this point. This information is generally necessary because the mathematical equations we use to describe physical systems are seldom exact, so that we only know the equilibrium points to a good approximation, or because we lack the ability to place the system precisely at the equilibrium point, or any combination of these and a host of other reasons.

D.2.1 Linearizing the System

To begin our investigation of the system near the equilibrium point, we *linearize* the system. This is a process by which the behavior of the nonlinear system is described in the vicinity of the equilibrium point through a linear approximation of the system.

Given the point $(\tilde{x}, \tilde{y})$, we make the definitions

$$\begin{aligned} x &= \tilde{x} + \xi \\ y &= \tilde{y} + \eta \end{aligned}$$

Taking the derivatives of these equations gives

$$\begin{aligned} \dot{x} &= \dot{\xi} = f(x, y) = f(\tilde{x} + \xi, \tilde{y} + \eta) \\ \dot{y} &= \dot{\eta} = g(x, y) = g(\tilde{x} + \xi, \tilde{y} + \eta) \end{aligned}$$

where the terms $\dot{\tilde{x}}$ and $\dot{\tilde{y}}$ do not exist because the values $\tilde{x}$ and $\tilde{y}$ are constants, and thus do not have derivatives.

We now use Taylor's Theorem to write the functions f and g as expansions about the equilibrium point (we ignore questions of differentiability of the functions at the equilibrium point, tacitly making whatever assumptions are necessary to complete the analysis). We have for f, for example,

$$\begin{aligned} f(x, y) &= f(\tilde{x} + \xi, \tilde{y} + \eta) \\ &= f(\tilde{x}, \tilde{y}) + \left.\frac{\partial f}{\partial x}\right|_{\tilde{x},\tilde{y}} \xi + \left.\frac{\partial f}{\partial y}\right|_{\tilde{x},\tilde{y}} \eta \\ &\quad + \frac{1}{2}\left[\left.\frac{\partial^2 f}{\partial x^2}\right|_{\tilde{x},\tilde{y}} \xi^2 + \left.\frac{\partial^2 f}{\partial y^2}\right|_{\tilde{x},\tilde{y}} \eta^2 + \left.\frac{\partial^2 f}{\partial x \partial y}\right|_{\tilde{x},\tilde{y}} \xi\eta \right] + \cdots \end{aligned}$$

Now, we are interested in the behavior of the system only near the equilibrium point, so we can safely assume that ξ and η are small. Since multiplying two small things together gives an even smaller result, we say that all *higher-order terms*, meaning terms in which ξ or η are squared, or multiplied by each other, are so small as to be negligible. Thus, we ignore all terms beyond the first derivative in the expansion we have just performed.

A second simplification can also be done. Recall that the definition of the point $(\tilde{x}, \tilde{y})$ is that $f(\tilde{x}, \tilde{y}) = 0$. Thus, another term drops out of the expansion, and when we plug it into our expression for $\dot{\xi}$ we have

$$\dot{\xi} = \left.\frac{\partial f}{\partial x}\right|_{\tilde{x},\tilde{y}} \xi + \left.\frac{\partial f}{\partial y}\right|_{\tilde{x},\tilde{y}} \eta$$

Similarly, the equation for $\dot{\eta}$ is

$$\dot{\eta} = \left.\frac{\partial g}{\partial x}\right|_{\tilde{x},\tilde{y}} \xi + \left.\frac{\partial g}{\partial y}\right|_{\tilde{x},\tilde{y}} \eta$$

The important thing about this formulation is that the derivatives are evaluated at the known point $(\tilde{x}, \tilde{y})$. This means that they are known constants, so we can write the differential equations for ξ and η in matrix form as

$$\left\{ \begin{array}{c} \dot{\xi} \\ \dot{\eta} \end{array} \right\} = \left[\begin{array}{cc} a & b \\ c & d \end{array} \right] \left\{ \begin{array}{c} \xi \\ \eta \end{array} \right\}$$

where

$$a = \left.\frac{\partial f}{\partial x}\right|_{\tilde{x},\tilde{y}} \qquad b = \left.\frac{\partial f}{\partial y}\right|_{\tilde{x},\tilde{y}} \qquad c = \left.\frac{\partial g}{\partial x}\right|_{\tilde{x},\tilde{y}} \qquad d = \left.\frac{\partial g}{\partial y}\right|_{\tilde{x},\tilde{y}}$$

and we write this in the generic form

$$\dot{\mathbf{r}} = A\mathbf{r} \tag{D.3}$$

D.2.2 Solution of the Linearized System

Linearizing the system about the equilibrium point has produced what is called a *constant-coefficient linear system of equations*. This is the vector version of the well-known scalar equation

$$\dot{x} = ax \tag{D.4}$$

The solution to (D.4) is $x(t) = \bar{x}e^{at}$, where $\bar{x}$ is the initial condition. Knowing this, we decide to try a solution to (D.3) of the form

$$\mathbf{r}(t) = \bar{\mathbf{r}}e^{\lambda t} \tag{D.5}$$

where λ is a constant to be determined. Plugging (D.5) into (D.3) gives

$$\begin{aligned}
\frac{d}{dt}(\bar{\mathbf{r}}e^{\lambda t}) &= A\bar{\mathbf{r}}e^{\lambda t} \\
\Rightarrow \lambda\bar{\mathbf{r}}e^{\lambda t} &= A\bar{\mathbf{r}}e^{\lambda t} \\
&\Rightarrow (A-\lambda I)\bar{\mathbf{r}}e^{\lambda t} = 0
\end{aligned} \tag{D.6}$$

Consider (D.6). We know that $e^{\lambda t} \neq 0$ for all values of λ and t, because the exponential is non-zero for all finite numbers, real or complex. Thus. if our equation is to be satisfied, we require

$$(A-\lambda I)\bar{\mathbf{r}} = 0 \tag{D.7}$$

Further, $\bar{\mathbf{r}} = 0$ is the *trivial* case, in which $\mathbf{r}(t) = 0$ for all t. This is a valid solution, but not an interesting one; it basically agrees that if the system is at zero to start with, it will stay there.

Now we recall something we learned in linear algebra class. That is,

Theorem D.1: Given that $M\mathbf{x} = 0$, where M is a square matrix and $\mathbf{x}$ is a vector of compatible dimension, either or both of the following must be true:

1. The vector $\mathbf{x} = 0$, or
2. The matrix M is singular; that is, the determinant of M is zero.

∎

We have already decided that we don't want to bother with $\bar{\mathbf{r}} = 0$, so that leaves the second possibility; we can satisfy (D.7) if the determinant $|A - \lambda I|$ is zero.

The values of λ for which $|A - \lambda I| = 0$ are known as the *eigenvalues* of the matrix A. We will not go into great detail about eigenvalues here, though they are of great import in the study of matrices, linear systems, and linear algebraic systems of all types. Instead, we simply note a couple of facts about eigenvalues that apply to linear systems of the type we are considering here.

1. A square matrix with n rows has exactly n eigenvalues.

2. Eigenvalues are in general complex.

The first of these facts means that for a system of two equations such as the one we are considering, there will be two solutions. The second means that the solutions will generally be described by the exponentials of complex numbers.

For a given eigenvalue, we can return to eqn. (D.7) and find a vector $\bar{\mathbf{r}}$ that satisfies the equation. Such a vector is known as the *eigenvector* corresponding to the given eigenvalue. Because the matrix $A - \lambda I$ is singular, $\bar{\mathbf{r}}$ is not quite unique; the values of the entries in the vector relative to each other are fixed, but the magnitude of the vector itself is arbitrary. To see this, consider some eigenvector multiplied by some non-zero scalar α, and plug this into eqn. (D.7):

$$(A - \lambda I)(\alpha\bar{\mathbf{r}}) = \alpha(A - \lambda I)\bar{\mathbf{r}}$$

Since we already know that $(A - \lambda I)\bar{\mathbf{r}} = 0$, this is satisfied for any value of α.

Since there are n eigenvalues, it follows that there are also n corresponding eigenvectors, and that each may have some arbitrary magnitude. Consider some general combination of the eigenvalue/eigenvector pairs, each with its scalar multiplier:

$$\mathbf{r}(t) = \alpha_1\bar{\mathbf{r}}_1 e^{\lambda_1 t} + \alpha_2\bar{\mathbf{r}}_2 e^{\lambda_2 t} + \cdots + \alpha_n\bar{\mathbf{r}}_n e^{\lambda_n t} \tag{D.8}$$

If we plug this back into the original equation (D.3), we have (after taking derivatives and re-arranging)

$$\alpha_1(\lambda_1\bar{\mathbf{r}}_1 - A\bar{\mathbf{r}}_1)e^{\lambda_1 t} + \alpha_2(\lambda_2\bar{\mathbf{r}}_2 - A\bar{\mathbf{r}}_2)e^{\lambda_2 t} + \cdots + \alpha_n(\lambda_n\bar{\mathbf{r}}_n - A\bar{\mathbf{r}}_n)e^{\lambda_n t} = 0 \tag{D.9}$$

We already know that each of the terms in parentheses is equal to zero, because each $\bar{\mathbf{r}}$ is a solution to the original differential equation. Thus eqn. (D.9) is satisfied, and the linear combination (D.8) is a solution to the differential equation (D.3).

D.2.3 Motion Near the Equilibrium Point

Consider now what the solution of our linear system means. We wrote the system as $\dot{\mathbf{r}} = A\mathbf{r}$, where

$$\mathbf{r} = \left\{ \begin{array}{c} \xi \\ \eta \end{array} \right\}$$

What this means is that $\mathbf{r}$ is the vector from the equilibrium point $(\tilde{x}, \tilde{y})$ to the nearby point $(\tilde{x} + \xi, \tilde{y} + \eta)$. We would like to know what happens as time passes; that is, how does $\mathbf{r}(t)$ change as $t \to \infty$?

Each possible solution, we said, is of the form $\mathbf{r}(t) = \bar{\mathbf{r}}e^{\lambda t}$, and any linear combination of these solutions is also a solution. It can also be shown that for a system of this type, all possible solutions are contained in the set of such solutions. Because this system is second-order, there are two solutions, and any possible solution can be written in the form

$$\mathbf{r}(t) = \alpha_1\bar{\mathbf{r}}_1 e^{\lambda_1 t} + \alpha_2\bar{\mathbf{r}}_2 e^{\lambda_2 t} \tag{D.10}$$

There are two terms on the right side of eqn. (D.10). Each term describes a particular *mode* of the system. The discussion of the system modes is beyond our needs here. It is sufficient to say that the general motion is a combination of the separate modes. The contribution of each mode to the motion changes with time, as shown my the exponential term. The values of the constants α_1 and α_2 are defined by the initial conditions. Since in general we cannot specify the initial conditions, we need to be concerned with all modes.

The analysis of any particular mode is identical to that of any other, so instead of worrying about which eigenvalue we are dealing with, we will use λ and it will be understood that any of the values of λ that cause (D.7) to be satisfied can be used.

Since in general λ may be complex, we write it as

$$\lambda = p + qi$$

where p and q are real and $i = \sqrt{-1}$. Then

$$\mathbf{r}(t) = \bar{\mathbf{r}} e^{\lambda t} = \bar{\mathbf{r}} e^{pt} e^{qit}$$

Now, e^{qit} is a complex term, and always has magnitude one in the complex plane. $\bar{\mathbf{r}}$ is constant, so it does not contribute to the behavior in any meaningful way. Thus, the term with which we are most concerned is e^{pt}.

D.2.4 Stability

By *stability*, we mean the property that, if we begin with (x, y) near the equilibrium point $(\tilde{x}, \tilde{y})$, it stays near the equilibrium point. For this simple two-dimensional system, this can be expressed mathematically as follows.

Definition D.1: The system

$$\begin{aligned} \dot{x} &= f(x, y) \\ \dot{y} &= g(x, y) \end{aligned}$$

is said to be *stable* at the point $(\tilde{x}, \tilde{y})$ if for every $\delta > 0$, there exists some $\varepsilon > 0$ such that

$$\|x(0) - \tilde{x}\| + \|y(0) - \tilde{y}\| < \varepsilon \quad \Longrightarrow \quad \|x(t) - \tilde{x}\| + \|y(t) - \tilde{y}\| < \delta$$

for all time $t > 0$. ♠

Note that the definition does not require that both x and y go back to $\tilde{x}$ and $\tilde{y}$. The extension of the definition to larger systems is obvious.

The reason for concern with e^{pt} is because it is the only term in the description of $\mathbf{r}(t)$ that may grow. Suppose $p > 0$. Then as $t \to \infty$, $pt \to \infty$. But the magnitude of $\mathbf{r}(t)$ is, we have argued, proportional to e^{pt}. And if pt gets large, the exponential of pt is not only going to get large, it is going to get large very quickly indeed.

But suppose $p < 0$. Then $t \to \infty$ causes $pt \to -\infty$, and therefore $e^{pt} \to 0$. That means that both ξ and η are approaching zero as time increases, which in turn means that the point $(x, y) = (\tilde{x} + \xi, \tilde{y} + \eta) \to (\tilde{x}, \tilde{y})$. What this means in English is that we started *near* the equilibrium point, and as time goes on we get closer to it.

The final possibility is that p might be identically zero. Then $e^{pt} = e^0 = 1$, and the vector $\mathbf{r}(t)$ is an oscillator with constant magnitude. This isn't as good as going to zero, but at least it means that $\mathbf{r}$ doesn't grow, so that the pair (x, y) stays in the vicinity of $(\tilde{x}, \tilde{y})$.

Since the general solution of the problem includes all the eigenvalues and eigenvectors, the only way we can be certain of stability is to have all of the eigenvalues satisfy the requirements. In general, we say that a linear system is *asymptotically* stable if all of the eigenvalues have negative real parts, *neutrally* stable if most of them have negative real parts and the others have zero real part, and *unstable* if *any* of the eigenvalues have positive real parts.

Note that since the system is normalized such that the equilibrium point is at $\mathbf{r} = 0$, asymptotic stability means that eventually the state of the system will converge to zero. Neutral stability does not make this guarantee; since some parts of the solution will oscillate forever, neutral stability only implies that the state of the system will not get arbitrarily large.

Example D.1. Consider the second-order system

$$\ddot{\theta} = k\cos\theta - c\dot{\theta} - w$$

(k, c, and w constants that describe the physical system). By letting

$$x = \theta, \qquad\qquad y = \dot{\theta},$$

this can be re-written as the two first-order equations

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= k\cos x - cy - w \end{aligned}$$

Find the equilibrium points of the system. Linearize the system and construct the eigenvalue equation. Then, taking the values $k = \sqrt{2}$, $w = 1$, show that the equilibrium points are $(\tilde{x}, \tilde{y}) = (-\pi/4, 0)$ and $(\pi/4, 0)$. Analyze these points, showing that $(\tilde{x}, \tilde{y}) = (-\pi/4, 0)$ is unstable for all values of c, and that $(\tilde{x}, \tilde{y}) = (\pi/4, 0)$ is stable for all $c \geq 0$.

Setting the velocities to zero gives

$$\dot{x} = 0 = \tilde{y} \quad \Longrightarrow \quad \tilde{y} = 0$$

as a requirement for any equilibrium point. Using this in the second equation gives

$$0 = k\cos\tilde{x} - w \quad \Longrightarrow \quad \cos\tilde{x} = w/k$$

Therefore the equilibrium points are any pair

$$(\tilde{x},\ \tilde{y}) = (\cos^{-1}(w/k),\ 0).$$

To linearize, write the system in the form

$$\begin{aligned} \dot{x} &= f(x, y) \\ \dot{y} &= g(x, y) \end{aligned}$$

and express the variables x and y as

$$x = \tilde{x} + \xi \qquad \text{and} \qquad y = \tilde{y} + \eta.$$

Then we expand about the equilibrium points, keeping only the first terms in the expansion (the 'linear' terms), to get

$$\left\{ \begin{array}{c} \dot{\xi} \\ \dot{\eta} \end{array} \right\} = \left[\begin{array}{cc} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \\ \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{array} \right]_{(\tilde{x},\tilde{y})} \left\{ \begin{array}{c} \xi \\ \eta \end{array} \right\}$$

In this case, the functions and their partial derivatives are

$$\begin{array}{rcl} f(x,y) & = & y \\ & & \partial f/\partial x = 0 \\ & & \partial f/\partial y = 1 \end{array} \qquad \begin{array}{rcl} g(x,y) & = & k\cos x - cy - w \\ & & \partial g/\partial x = -k\sin x \\ & & \partial g/\partial y = -c \end{array}$$

The eigenvalue equation is then

$$|\lambda I - A| = \left| \left[\begin{array}{cc} \lambda & 0 \\ 0 & \lambda \end{array} \right] - \left[\begin{array}{cc} 0 & 1 \\ -k\sin\tilde{x} & -c \end{array} \right] \right| = \lambda(\lambda + c) + k\sin\tilde{x} = 0$$

For the given values of k and w, the equation for $\dot{y}$ (recalling $\tilde{y}$ must always be zero) becomes

$$\dot{y} = \sqrt{2}\cos\tilde{x} - 1 = 0 \quad \Longrightarrow \quad \cos\tilde{x} = \sqrt{2}/2$$

The possible values for $\tilde{x}$ are $\pm\pi/4$, as requested in the problem statement.

Consider the equilibrium point $(-\pi/4,\ 0)$. The eigenvalue equation becomes

$$\lambda^2 + c\lambda + \sqrt{2}\sin(-\pi/4) = \lambda^2 + c\lambda - 1 = 0.$$

Solving gives

$$\lambda = \frac{1}{2}\left[-c \pm \sqrt{c^2+4}\,\right].$$

Considering the term under the radical, we see that it is always positive, and that

$$c^2 + 4 > c^2 \quad \Longrightarrow \quad \sqrt{c^2+4} > |c|.$$

Since the value under the radical is positive, both of the eigenvalues are real. Due to the magnitude of the term under the radical, one of the eigenvalues is certain to be positive, regardless of the value of c. Therefore, since a positive real part to an eigenvalue means that the system is unstable, this is an unstable equilibrium point.

For the equilibrium point $(\tilde{x},\tilde{y}) = (\pi/4, 0)$, we have the eigenvalue equation for this equilibrium point to be

$$\lambda^2 + c\lambda + 1 = 0$$

from which the eigenvalues are

$$\lambda = \frac{1}{2}\left[-c \pm \sqrt{c^2-4}\right].$$

The system is unstable if there is a value of c for which one of the two roots is real and positive. Consider the three possibilities

$$\begin{array}{ll} \text{(a)} & c^2 < 4 \\ \text{(b)} & c^2 \geq 4 \\ \text{(c)} & c = 0 \end{array}$$

In the first case, the term under the radical is negative, so the sign of the real part depends only on the sign of c. This will be negative if $c > 0$, so at least for $c < 2$, the system is stable for positive c.

In the second case, the term under the radical is positive (or zero), and has magnitude less than c^2. Therefore, both roots will be real and have the same sign, and will be negative if $c > 0$.

Finally, in the third case, the roots are $\lambda = \pm i$, which is purely imaginary. In this case, the system is neutrally stable.

Summarizing these results, the system is stable if $c \geq 0$. ♠

D.3 Second-Derivative Systems

The system we looked at in the last several pages was described by a pair of *first-order* equations; that is, there were only first derivatives with respect to time involved. For various reasons, we will want to examine systems of *second*-order equations; in particular, the equations of the three-body problem will be analyzed in their second-order form.

In systems described by second-order equations, we have both velocity and acceleration for each of the state variables. With the two states x and y, this leads to the system

$$\ddot{x} = f(x, \dot{x}, y, \dot{y}) \tag{D.11}$$

$$\ddot{y} = g(x, \dot{x}, y, \dot{y}) \tag{D.12}$$

An equilibrium point again requires that the states of the system be constant. Therefore the velocities and the accelerations will be zero. [1]

Assume as before that such a point is found. As in the last section, we consider motion near the point by expanding about it, but in this case we allow for small velocities as well as small displacements. Again using ξ and η, we have

$$\begin{array}{lclcll} x & = & \tilde{x} + \xi & \Longrightarrow & \dot{x} = \dot{\xi}; & \ddot{x} = \ddot{\xi} \\ y & = & \tilde{y} + \eta & \Longrightarrow & \dot{y} = \dot{\eta}; & \ddot{y} = \ddot{\eta} \end{array}$$

[1] In many applications, it is enough to find a point of *pseudo-equilibrium*, at which only the accelerations are required to be zero. This will not be the case in the applications we consider.

Proceeding as for systems of first-order equations, we take the Taylor expansion about the equilibrium point, drop the higher-order terms, and eliminate the values that we know are zero. The only complication is that we must take partial derivatives with respect to the velocities as well as the positions. This is not difficult, however. When put into matrix form, the resulting equations are

$$\left\{ \begin{array}{c} \ddot{\xi} \\ \ddot{\eta} \end{array} \right\} = \left[\begin{array}{cc} \frac{\partial f}{\partial \dot{x}} & \frac{\partial f}{\partial \dot{y}} \\ \frac{\partial g}{\partial \dot{x}} & \frac{\partial g}{\partial \dot{y}} \end{array} \right]_{\tilde{x},\tilde{y}} \left\{ \begin{array}{c} \dot{\xi} \\ \dot{\eta} \end{array} \right\} + \left[\begin{array}{cc} \frac{\partial f}{\partial \dot{x}} & \frac{\partial f}{\partial \dot{y}} \\ \frac{\partial g}{\partial \dot{x}} & \frac{\partial g}{\partial \dot{y}} \end{array} \right]_{\tilde{x},\tilde{y}} \left\{ \begin{array}{c} \xi \\ \eta \end{array} \right\} \tag{D.13}$$

We are now looking at systems of the form

$$\ddot{\mathbf{r}} = A\dot{\mathbf{r}} + B\mathbf{r} \tag{D.14}$$

Again, we will assume a solution of the form $\mathbf{r}(t) = \bar{\mathbf{r}}e^{\lambda t}$ Plugging this into (D.14), we have

$$\begin{aligned} \lambda^2 \bar{\mathbf{r}} e^{\lambda t} &= \lambda A \bar{\mathbf{r}} e^{\lambda t} + B \bar{\mathbf{r}} e^{\lambda t} \\ &\Rightarrow (\lambda^2 I - A\lambda - B)\bar{\mathbf{r}} = 0 \end{aligned}$$

Again, we are interested in the values of λ for which the determinant of the matrix is zero;

$$|\lambda^2 I - A\lambda - B| = 0 \tag{D.15}$$

In the case we looked at previously, we said that for a system of two equations, there would always be two eigenvalues. In this case, the system may still have only two equations, but each equation is second order. That means that each equation contributes two eigenvalues, and there are a total of four values of λ for which (D.15) is satisfied.

Since the solutions are of the same form as those for the first-derivative system, the stability analysis is identical. Again, we arrive at the requirement that the eigenvalues all have non-positive real parts for stability.

D.3.1 Motion near the Equilibrium Point

As with the system of first-order equations, the response of the second derivative system can be described as a linear combination, as in (D.8). Recall that each of the constant vectors describes a particular mode of motion. In this case, the modes will include both the states ξ and η and their derivatives. We include all of these in an augmented state vector, which we will denote $\mathbf{R}$ to avoid confusion. We have

$$\mathbf{R} = \left\{ \begin{array}{c} \mathbf{r} \\ \dot{\mathbf{r}} \end{array} \right\} = \left\{ \begin{array}{c} \xi \\ \eta \\ \dot{\xi} \\ \dot{\eta} \end{array} \right\}$$

Using this definition in (D.14), we have a system of four first-derivative equations written as

$$\dot{\mathbf{R}} = \begin{bmatrix} 0 & I \\ B & A \end{bmatrix} \mathbf{R} \tag{D.16}$$

The matrix in this equation is 4 by 4, and thus has four eigenvalues. It can be shown that the eigenvalues are the same as those computed from (D.15). Further, the eignevectors associated with this system are those for the second-derivative system in its original form. Note that in general there will be *four* such eigenvectors for this system. This reflects the fact that there are four variables (ξ, η, $\dot{\xi}$, and $\dot{\eta}$) for which initial values may separately be specified.

D.3.2 Stability of Non-dissipative Systems

One final comment has to be made, this one concerning systems in which there are no friction or damping forces. Such forces are known as *dissipative*, because they dissipate energy from the system. To achieve asymptotic stability for a linear system, some such dissipative mechanism must exist.

Remember that for asymptotic stability, the state of the system goes to zero. As an example, this might relate to a pendulum ceasing its oscillations and hanging still. In this case, both the potential energy and the kinetic have gone to zero. The primary dissipative mechanism is air resistance.

If there is no dissipation, the total energy of the system can never decrease. What this means in practice is that the system either has kinetic energy, so that it is moving, or it has potential energy, and thus is about to move, or both.

Orbital systems, in the idealized analysis, in general have no dissipative mechanism. In the restricted three-body problem, for instance, there is no dissipative term in the equations for motion near the equilibrium point. Therefore, there is no way to get asymptotic stability. The mathematical result of this is that the eigenvalues will not have negative real parts, which would imply such stability. Instead, the best that can be hoped for is neutral stability, with all eigenvalues having zero real part.

Appendix E

Tables

This appendix contains lists of physical constants and planetary information. The tables in various textbooks and sources disagree, often in the fourth or fifth decimal place, sometimes in the third. The values given here are representative, and are the ones used in the examples in the text.

Some of the information in planetary tables may seem inconsistent. For example, the period of a planet's orbit calculated from the tabulated data may not agree precisely with the tabulated period. This will be due to both the rounding of the values listed in the table, and the fact that the two-body equations derived in the text do not allow for many very small perturbations that occur in fact. These are too small to be included in the calculations (and are unnecessary for our purposes), but are large enough to appear in the precisely measured values in the tables. The tables included here do not have the period listed; when needed, we will assume that the two-body calculation is sufficiently close.

The sources of this information include several of the listed references. Also, some is taken from on-line sources such as:

1) *Wikipedia – A Free Encyclopedia* at `http://www.wikipedia.org`

E.1 Basic Solar System Units

$\mu_\odot$	Gravitational parameter of the Sun	$1.32715 \times 10^{11}\,\mathrm{km}^3/\mathrm{s}^2$
AU	Astronomical unit	$1.495979 \times 10^{8}\,\mathrm{km}$

Note: The astronomical unit is no longer defined as the semimajor axis of the Earth's orbit. It is now defined as the radius of the orbit of a particle of negligible mass about the Sun, with period one Gaussian year (365.2568983 mean solar days; this was the value calculated by Gauss as the length of the sidereal year). The difference appears in the sixth significant digit.

E.2 Earth Physical Data

$\mu_\oplus$	Gravitational parameter	$3.98601 \times 10^{5}\,\mathrm{km}^3/\mathrm{s}^2$
$r_\oplus$	Equatorial radius	6378.145 km
	Inclination of equator	23.45°
	Sidereal year	365.256 mean solar days
	Mean solar day	86400 mean solar seconds
	Sidereal day	$23^{\mathrm{h}}\,56^{\mathrm{m}}\,04^{\mathrm{s}}$ mean solar time

Notes:

1. The polar radius of the Earth is slightly (about 20 km) less than the equatorial.

2. In addition to the mean solar second (defined as 1/86400 mean solar days), there is also a sidereal second, which is 1/86400 sidereal days. One mean solar second = 1.0027379093 sidereal seconds.

E.3 Planetary Data

E.3.1 Orbital Characteristics

The following data is taken from Danby [6], with a couple of corrections from Vallado [16]. With the exception of the data for Pluto, It is referenced to the epoch January 1, 2000. Pluto data is taken from the *Astronomical Almanac* for 1995.

	Planet	semi-major axis (AU)	eccentricity	inclination (degrees)
☿	Mercury	0.3871	0.2056	7.005
♀	Venus	0.7233	0.0068	3.395
⊕	Earth	1.0000	0.0167	0.000
♂	Mars	1.5237	0.0934	1.850
♃	Jupiter	5.2026	0.0485	1.303
♄	Saturn	9.5381	0.0555	2.489
⛢	Uranus	19.1833	0.0463	0.773
♆	Neptune	30.0551	0.0090	1.770
♇	Pluto	39.5376	0.2509	17.13

E.3.2 Physical Data

	Planet	Equatorial radius (km)	Mass (Earth = 1)
☿	Mercury	2439	0.056
♀	Venus	6051	0.817
⊕	Earth	6378	1.000
♂	Mars	3393	0.108
♃	Jupiter	71400	318.0
♄	Saturn	60000	95.2
⛢	Uranus	25400	14.6
♆	Neptune	25295	17.3
♇	Pluto	1500	0.0022

E.4 The Moon

$r_☾$	Radius	1738 km
$a_☾$	Semimajor axis	384400 km
	Mass	0.0123 × Earth mass

Index